

Tabelas de contingência: teste qui-quadrado de Pearson, teste exato de Fisher, odds ratio e V de Cramer

*Paulo S. P Silveira (paulo.silveira@fm.usp.br)
Koichi Sameshima (koichi.sameshima@fm.usp.br)
José O. Siqueira (jose.siqueira@fm.usp.br)*

Contents

Objetivos	2
Preparação	2
Contingência	2
Relações entre duas variáveis	3
tabela 2x2	3
independência	4
concordância	5
discordância	6
Associação	6
causa e efeito positivo?	7
causa e efeito negativo?	8
causa e efeito positivo?	9
Exemplos	10
Gatos e alergia	10
- estresse	11
- mecanismo imune	12
- desinfetantes	13
- alergia e estresse	14
- alergia e gatos	15
Vinho e doença coronariana	16
- ambiente inseguro	17
- moradia precária	18
- má nutrição	19
- boa moradia	20
- segurança	21
- boa alimentação	22
- outros cuidados	23
- é o vinho?	24
Fumo e Câncer de Pulmão	25
- The Cigarette Century	26
- How could one attribute lung cancer to cigarette smoking?	26
- Medical knowledge is provisional	26
- Fisher was against the association between smoking and lung cancer	26
Óculos e xixi	28

A hipótese nula	29
Teste qui-quadrado de Pearson para Independência	29
- a distribuição χ^2	30
Exemplo: capacete e trauma	36
- decisão pela estatística de teste	41
- decisão pelo valor-p	41
Teste do χ^2 robusto	42
Riscos relativos	42
razão de riscos (RR, <i>risk ratio</i>)	43
razão de chances (OR, <i>odds ratio</i>)	44
- decisão pelo intervalo de confiança	45
Dois códigos úteis, com vários dos elementos apresentados neste texto	46
tabelacontingencia2x2_capacetetrauma.R	46
Tabelas LxC	51
Métodos tradicionais	52
v20200325.1343	

Objetivos

- Identificar e sumariar dados categóricos em tabela de contingência.
- Definir e determinar frequências observada e esperada.
- Formular hipótese estatística na análise de dados categóricos.
- Identificar, aplicar e interpretar os testes de proporção.
- Identificar distribuição amostral χ^2 (qui-quadrado).
- Identificar e formular análise de variáveis categóricas em tabela de contingência *lxc* (números de linhas versus de colunas).
- Realizar os procedimentos estatísticos em R.

Preparação

Para executar os exemplos aqui apresentados, sugerimos que crie um projeto e baixe os seguintes arquivos para a pasta do mesmo:

- “tabelacontingencia2x2_capacetetrauma.R”
- “Qui2.R”
- “Qui2_LxC.R”
- “friendlycolor.R”

Contingência

Em sentido geral, contingência pode significar qualquer relação de dependência entre eventos ambientais ou entre eventos comportamentais e ambientais (Catania, 1993; Skinner, 1953, 1969; Todorov, 1985). Embora possa ser encontrado nos dicionários com diferentes significados, esse termo é empregado, na análise do comportamento, **como termo técnico para enfatizar como a probabilidade de um evento pode ser afetada ou causada por outros eventos** (Catania, 1993, p. 368).

Relações entre duas variáveis

1 Exposição e Desfecho

tabela 2x2

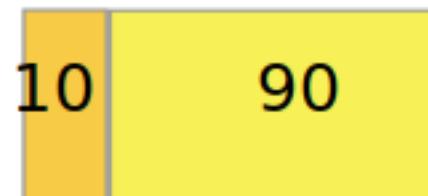
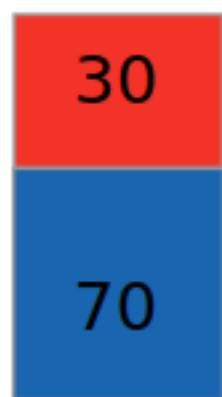
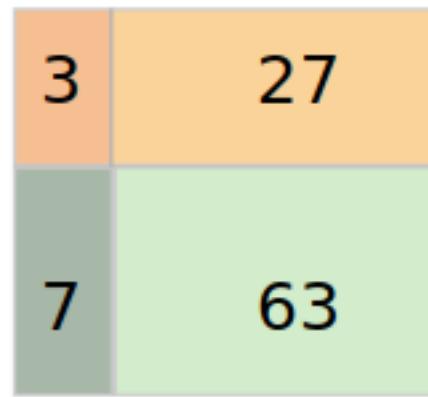
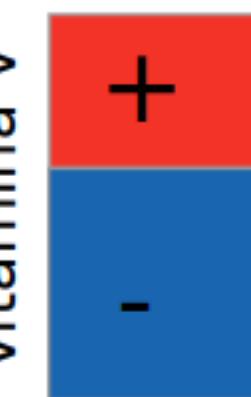
		Desfecho	
		+	-
Fator de exposição	+	+	-
	-	-	-

independência

Acordar cedo

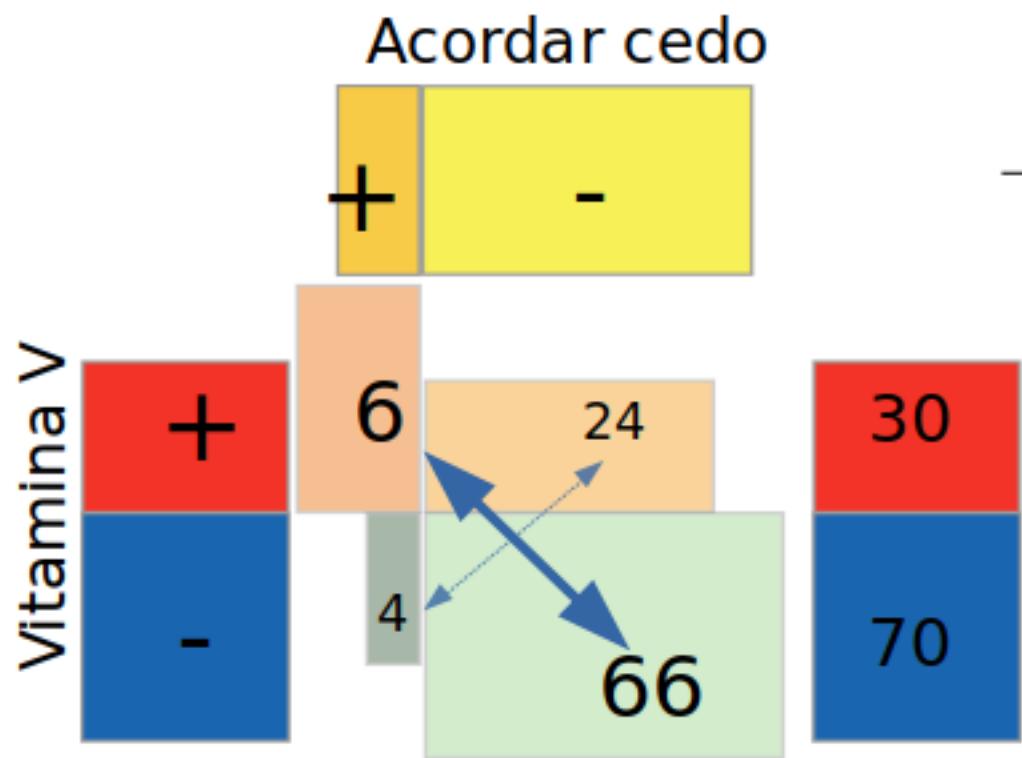


Vitamina V



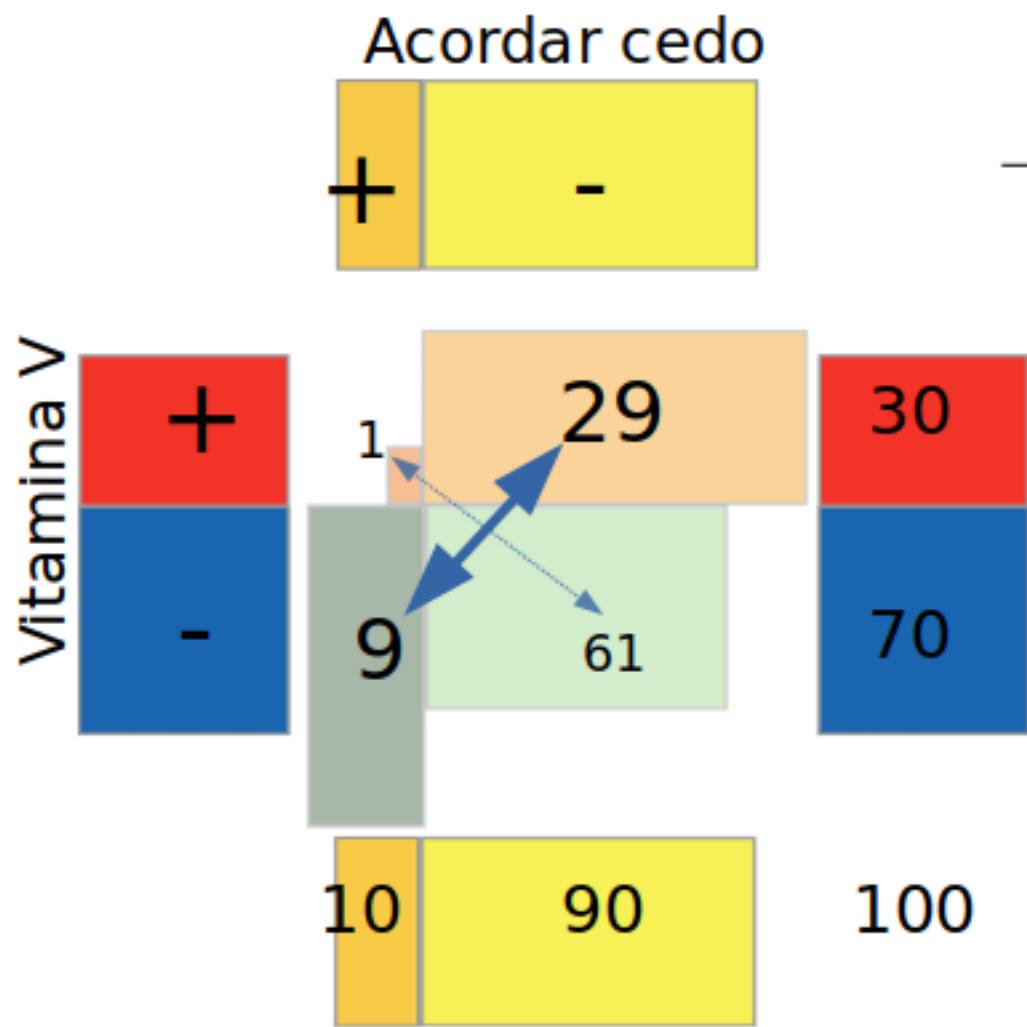
100

concordância



▷

discordância



Associação

2 Não avalia causa e efeito: apenas associação.

causa e efeito positivo?

Causa e efeito +?



causa e efeito negativo?

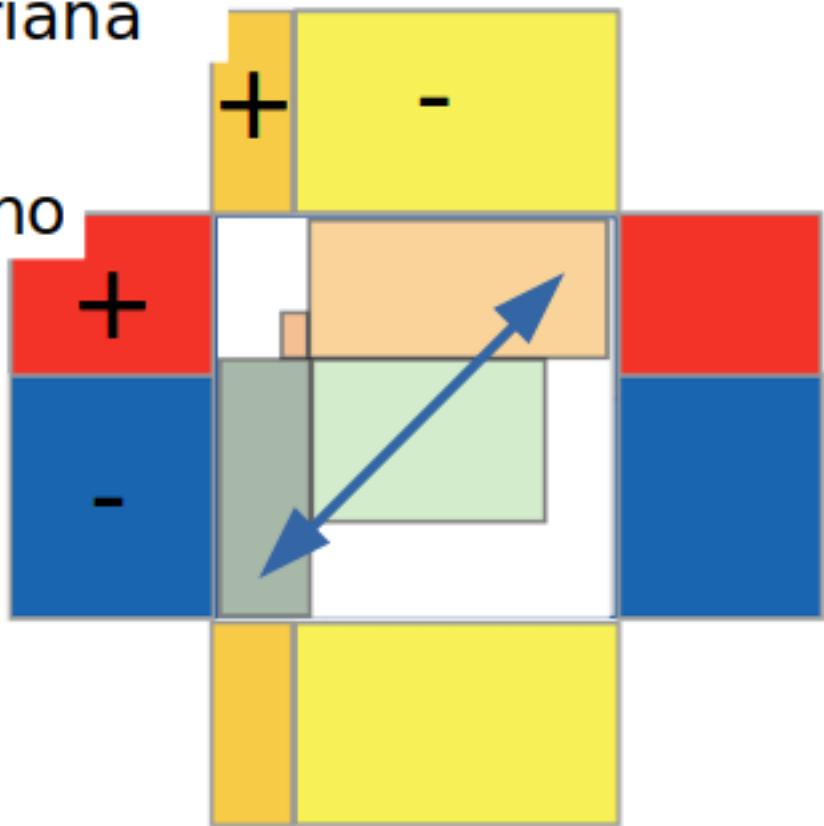
Causa e efeito -?



Reduz doença
coronariana

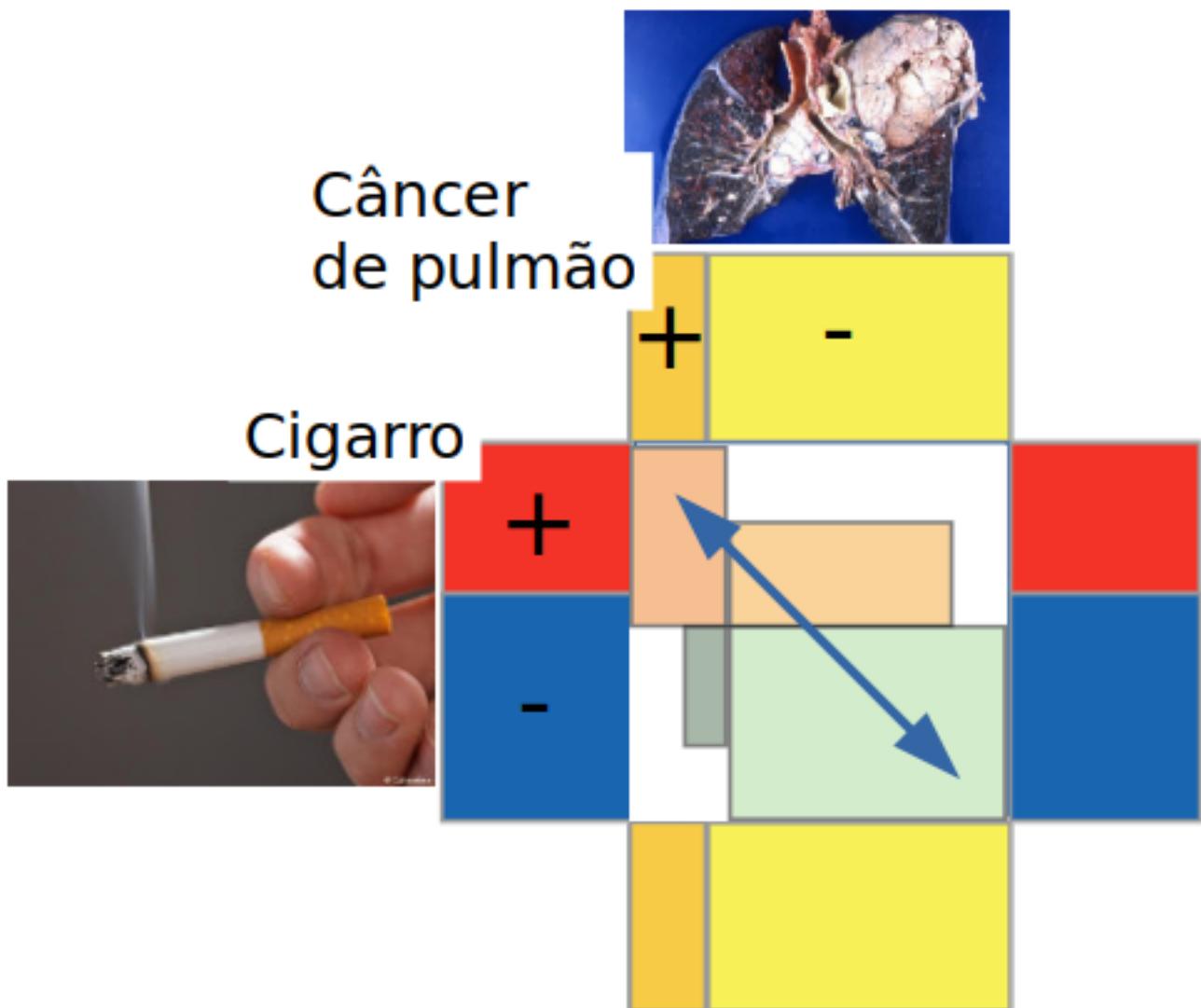


Vinho



causa e efeito positivo?

Causa e efeito +?



Exemplos

Gatos e alergia

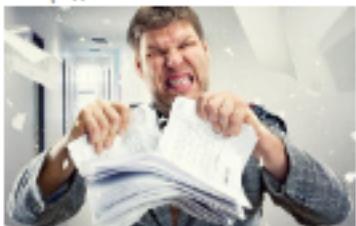
Causa e efeito +?



- estresse

Causa e efeito +?

<http://www.sbie.com.br>



Estresse



Alergia



Gatos



- mecanismo imune

Causa e efeito +?



Estresse



Gatos



<https://www.solucoesindustriais.com.br>



Imune

Alergia



- desinfetantes

Causa e efeito +?

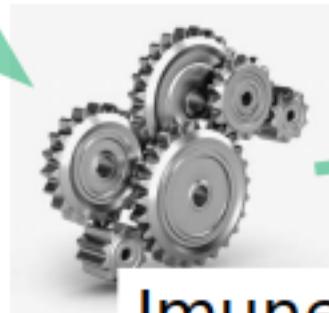


Estresse



Gatos

Desinfetantes



Imune



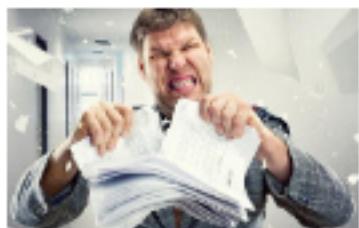
Alergia



Ácaros

- alergia e estresse

Causa e efeito +?



Estresse



Gatos

Desinfetantes



Imune



Alergia

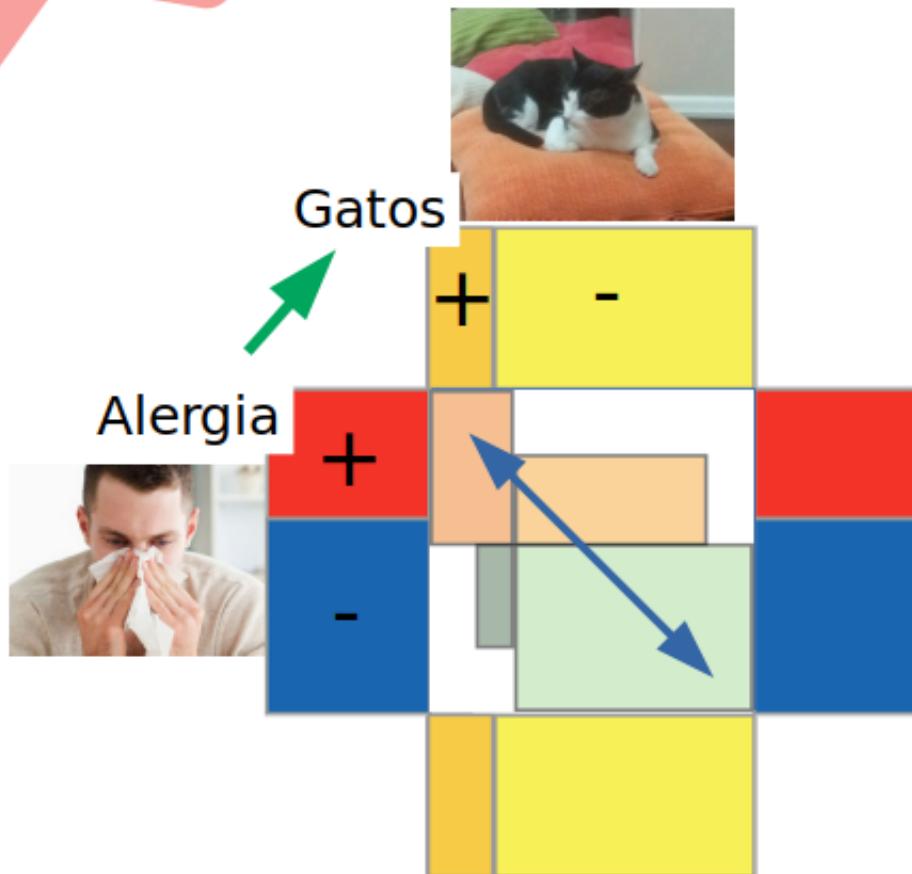


Ácaros

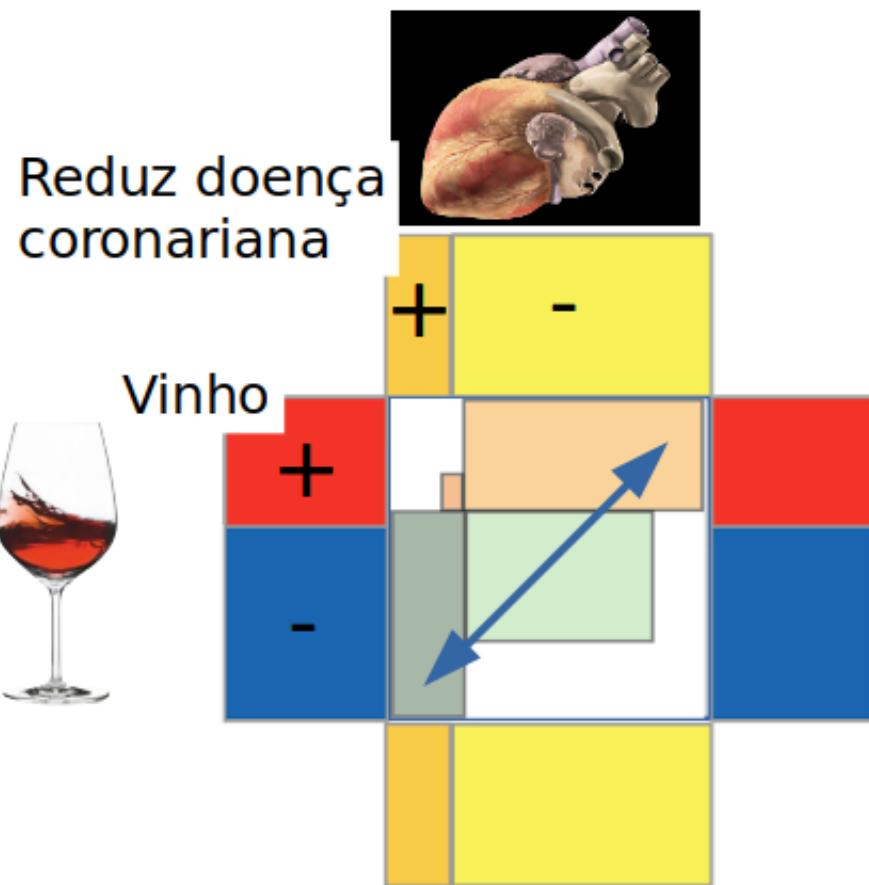
- alergia e gatos

Causa e efeito

Associação (sem direção, sem mecanismo)



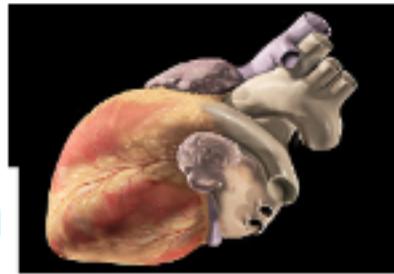
Associação



- ambiente inseguro

Associação

Reduc doença
coronariana ?



- moradia precária

Associação

Reduc doença
coronariana ?



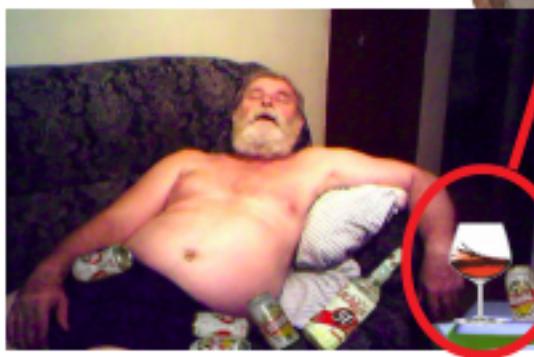
<https://www.campograndenews.com.br>



- má nutrição

Associação

Reduc doença
coronariana ?



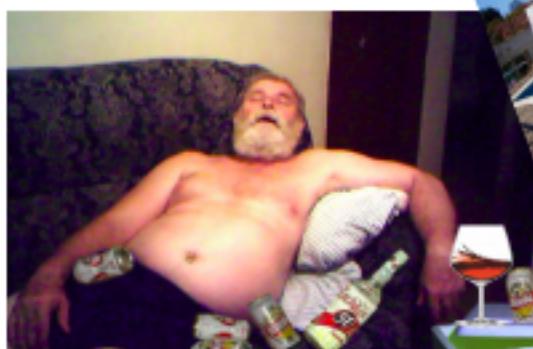
<https://www.solbrilhando.com.br>



- boa moradia

Associação

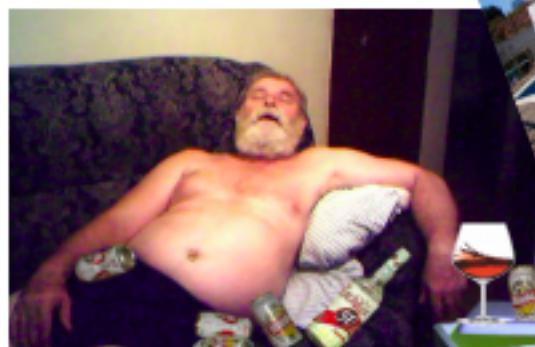
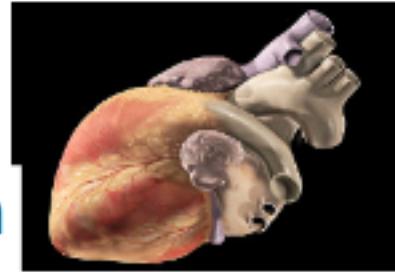
Reduc doença
coronariana ?



- segurança

Associação

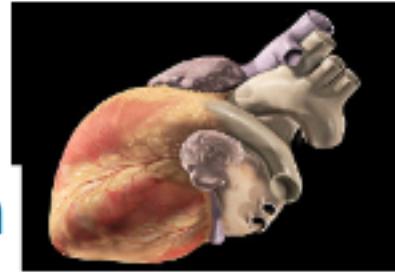
Reduc doença
coronariana ?



- boa alimentação

Associação

Reduc doença
coronariana ?



<https://exame.abril.com.br>

- outros cuidados

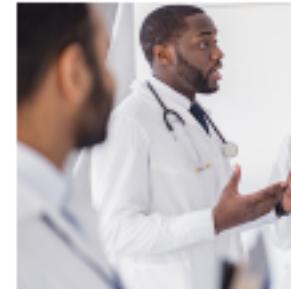
Associação



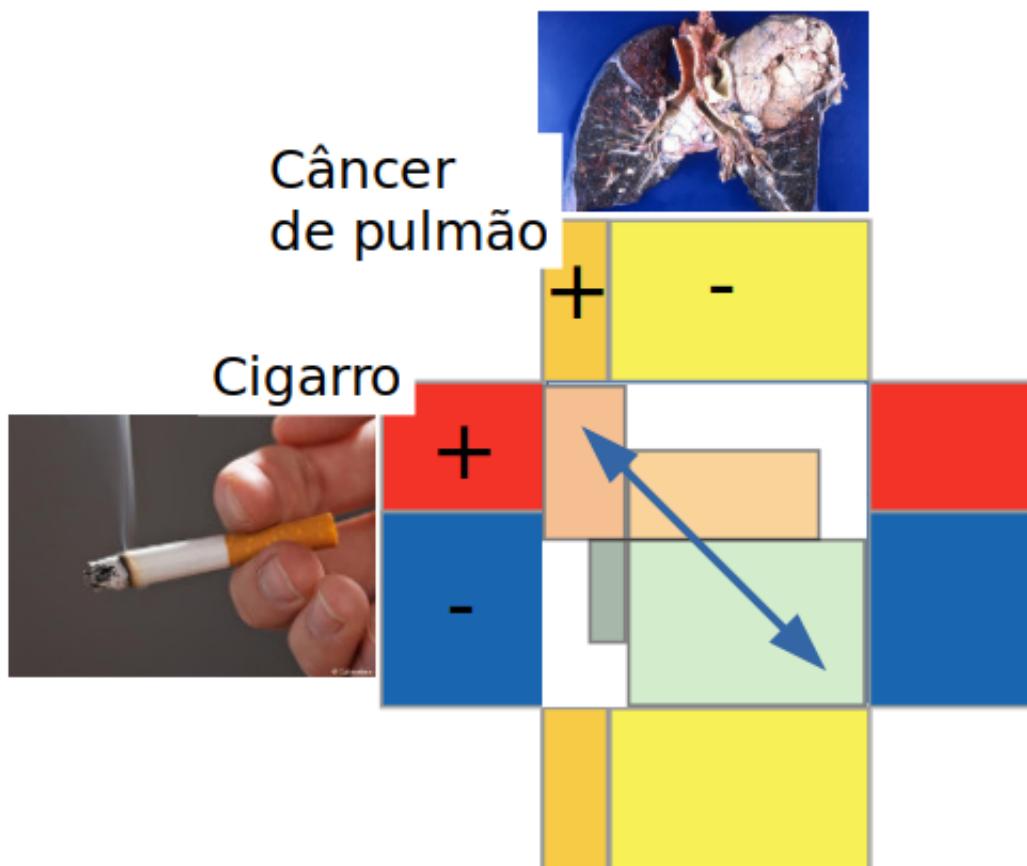
- é o vinho?

Associação

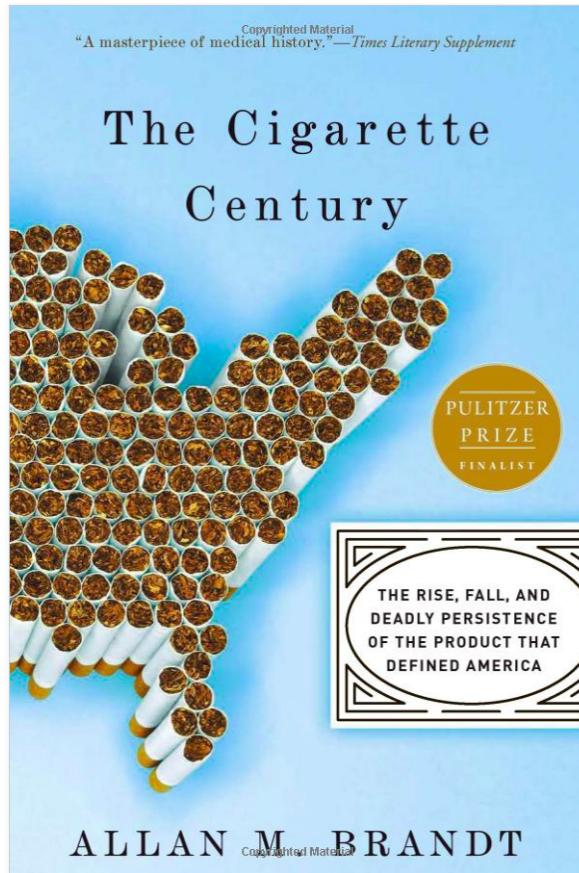
Reduc doença
coronariana ?



Causa e efeito +?



- The Cigarette Century



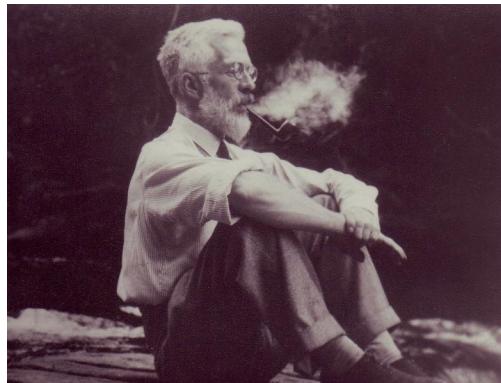
- How could one attribute lung cancer to cigarette smoking?

"The public impression of scientific and medical **uncertainty** that resulted was a crucial element in maintaining the market of current smokers as well as recruiting new ones. Industry literature, for example, frequently pointed to the fact that **nonsmokers could also develop lung cancer**. Therefore, they argued, **how could one attribute lung cancer to cigarette smoking?**"

- Medical knowledge is provisional

medical knowledge was always provisional and **contingent**. Just as drugs deemed "effective" do not work in every case, so too **a cause of disease does not always result in disease**. As Richard Doll would later explain, the epidemiologists had identified a cause of lung cancer (and other diseases), not the cause.

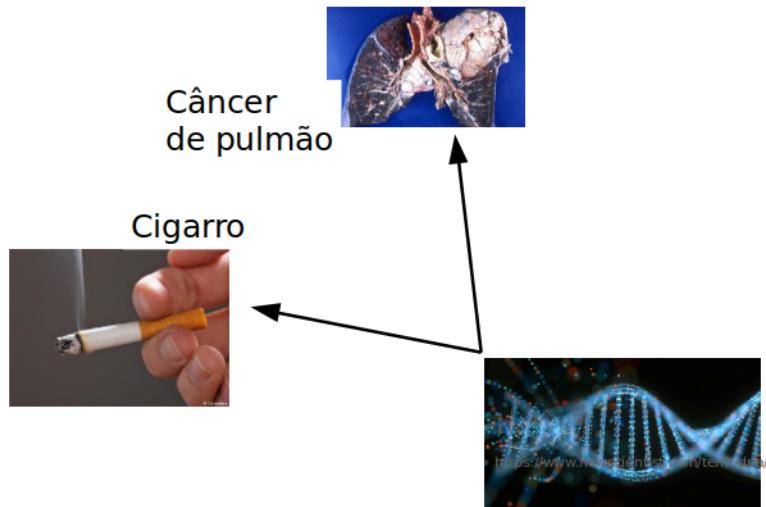
- Fisher was against the association between smoking and lung cancer



Another vocal critic of the lung cancer findings was **Sir Ronald Fisher**, the leading biometristian and geneticist in Great Britain during the first half of the twentieth century and a man deeply committed to bringing statistical analysis to genetics and agricultural experimentation. His **1925** book, Statistical Methods for Research Workers, quickly became a classic, leading to academic appointments at University College London and Cambridge University. Fisher's critiques were similar to Berkson's. The ethical impossibility of conducting a randomized experiment led him to question the results of the epidemiological studies. As a **believer in genetic* notions of cancer causality**, Fisher speculated that there was some constitutional factor that led individuals both to become smokers and to get lung cancer, even though smoking and lung cancer might not be causally related. Doll and Hill repeatedly rebutted this theory, returning to the critical question of how to account for the rise in lung cancers during the twentieth century if the disease was simply "constitutional."

* **1923:** One of the [A. Bradford] Hill's innovations was the **first randomized, double-blind clinical trial**, designed to reduce investigator bias in the evaluation of clinical outcomes. [...] This method, which drew on Fisher's **agricultural experimentation** in genetics, became a critical new tool for evaluating medical treatments.

Causa e efeito +?

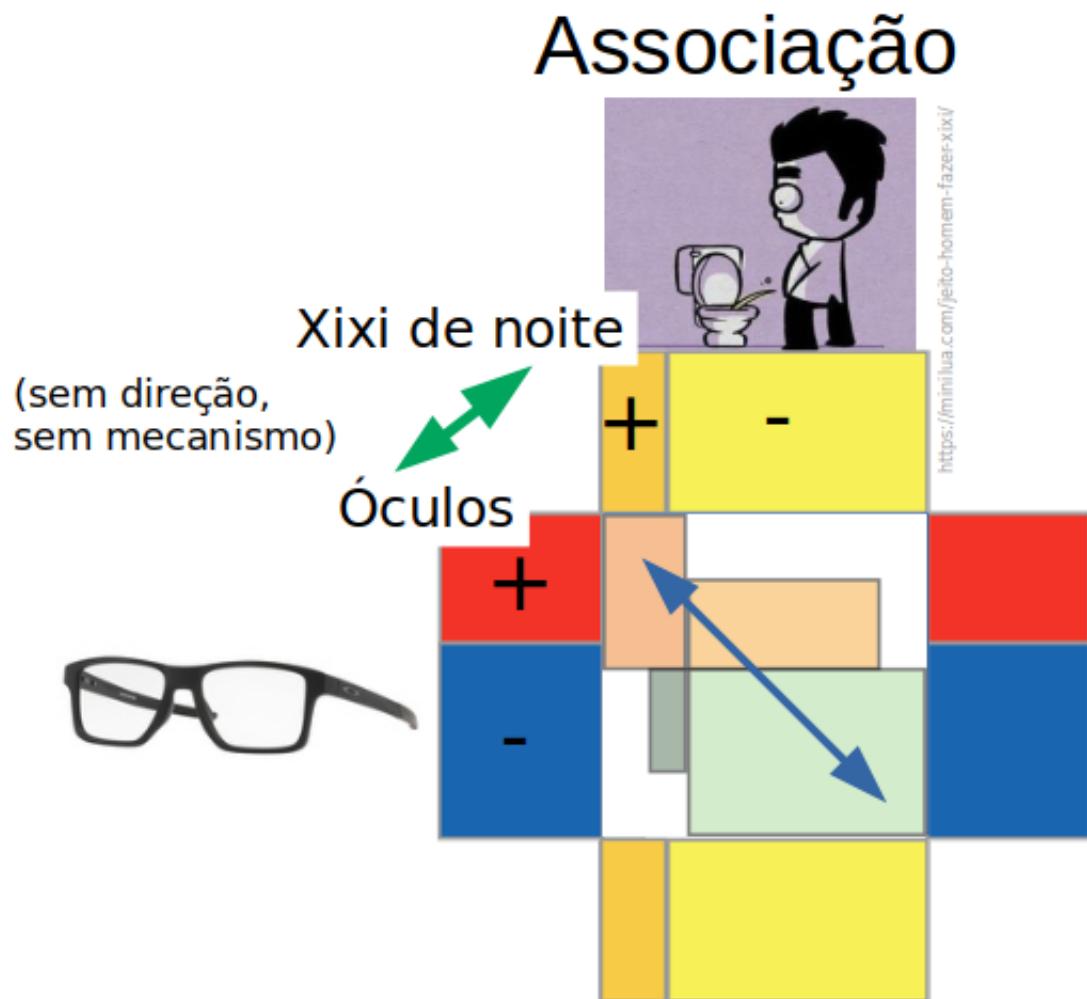


By the **1930s**, the relationship of smoking to cancer was a topic of unresolved debate. It was the **life insurance industry**, which like the tobacco corporations grew by leaps and bounds in the first half of the twentieth century, that **took the lead in understanding the effects of smoking on health**.

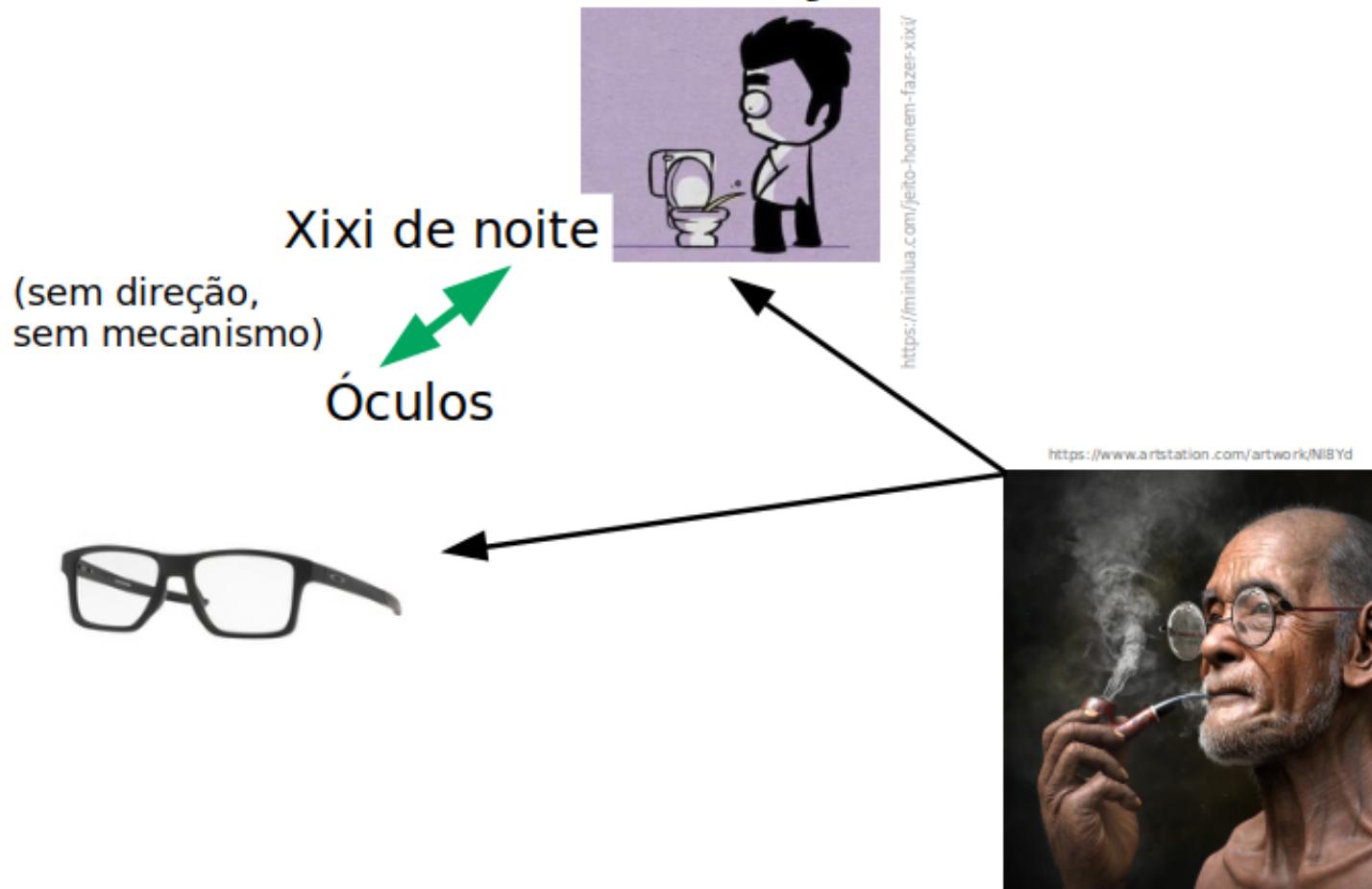
By the early **1950s**, however, it was abundantly clear that the evidence implicating cigarette smoking as a risk to health was now of a different order. First, the link between smoking and disease was **categorical**, outside the realm of individual clinical judgment.

They [tobacco industry] responded with a new and unprecedent public relations strategy. Its goals was to produce and **sustain scientific skepticism and controversy** in order to disrupt the emerging consensus on the harms of cigarette smoking. This strategy required **intrusions into scientific process and procedure**. [...] So long as there appeared to be doubt, so long as the industry could assert “**not proven**”. [...] The future of the cigarette would now depend on the **successful production of a scientific controversy**.

Óculos e xixi



Associação



Quando alguém lhe traz uma opinião coincidente com sua crença, você a toma como um fato.
Quando alguém lhe traz um fato discordante de sua crença, você o toma como uma opinião.

Prof. Eduardo Massad

A hipótese nula

3 Tomada de decisão: Hipótese nula (H_0) e hipótese alternativa (H_1)

Teste qui-quadrado de Pearson para Independência

O teste qui-quadrado de Pearson permite que se descubra se existe um relacionamento ou efeito de interação entre duas variáveis qualitativas nominais com duas ou mais categorias.

Cada observação da unidade observacional deve ser alocada em apenas uma categoria de cada uma das duas variáveis nominais.

SIEGEL, S. & CASTELLAN Jr., N.J. (1988) Nonparametric statistics for the behavioral sciences. 2^a ed. NY: McGraw-Hill, p. 111

- a distribuição χ^2

O comando para calcular a probabilidade acumulada entre zero e cada valor de qui-quadrado é:

`dchisq(x, df, ncp = 0, log = FALSE)`

onde x é o valor de χ^2 , df são os graus de liberdade (*degrees of freedom*) e ncp é o parâmetro de não centralidade (*non-centrality parameter* = 0, por *default*, correspondente à distribuição centrada de H_0 ; voltaremos ao ncp adiante). O parâmetro log, desligado por *default* é para a possibilidade de devolver o logarítmico do valor (não o exploraremos neste texto).



Os graus de liberdade em uma tabela com L linhas e C colunas é dado por

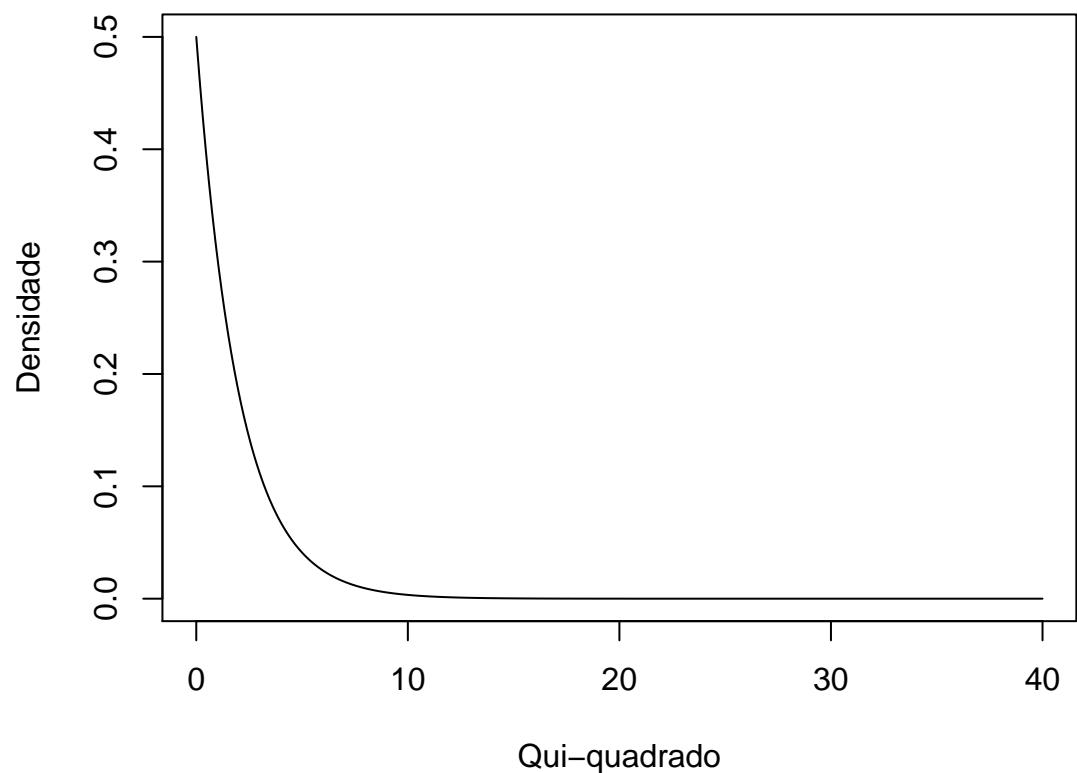
$$df = (L - 1) \cdot (C - 1)$$

Portanto, uma tabela 2 x 2, tem somente um grau de liberdade, e uma tabela 4 x 3 tem seis graus de liberdade.

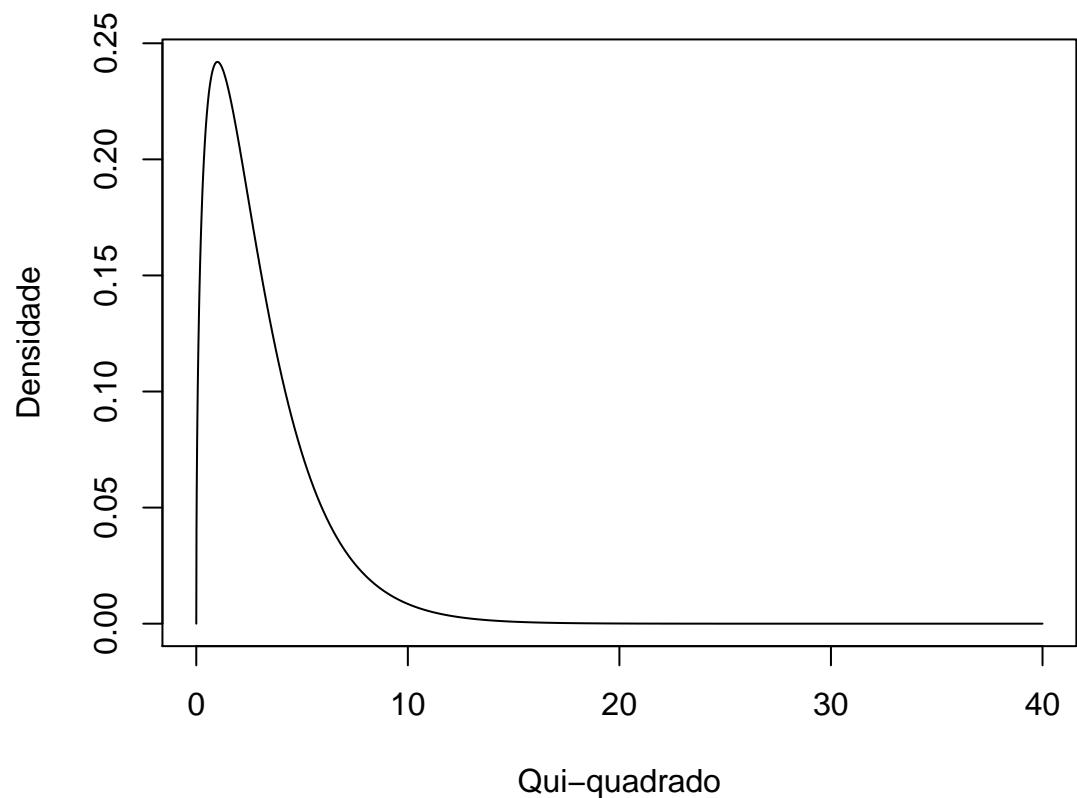
Pode-se gerar curvas desta distribuição com o seguinte código, mostrando seu aspecto entre 2 e 6 graus de liberdade:

```
chi2.valor <- seq(from=0, to=40, by=0.01)
graus.liberdade <- 2:6
for (g in 1:length(graus.liberdade))
{
  p <- dchisq(chi2.valor, df=graus.liberdade[g])
  plot(chi2.valor, p, type="l",
       main=paste("Distribuição de Qui^2, df=",
                  graus.liberdade[g], sep=""),
       xlab="Qui-quadrado", ylab="Densidade")
}
```

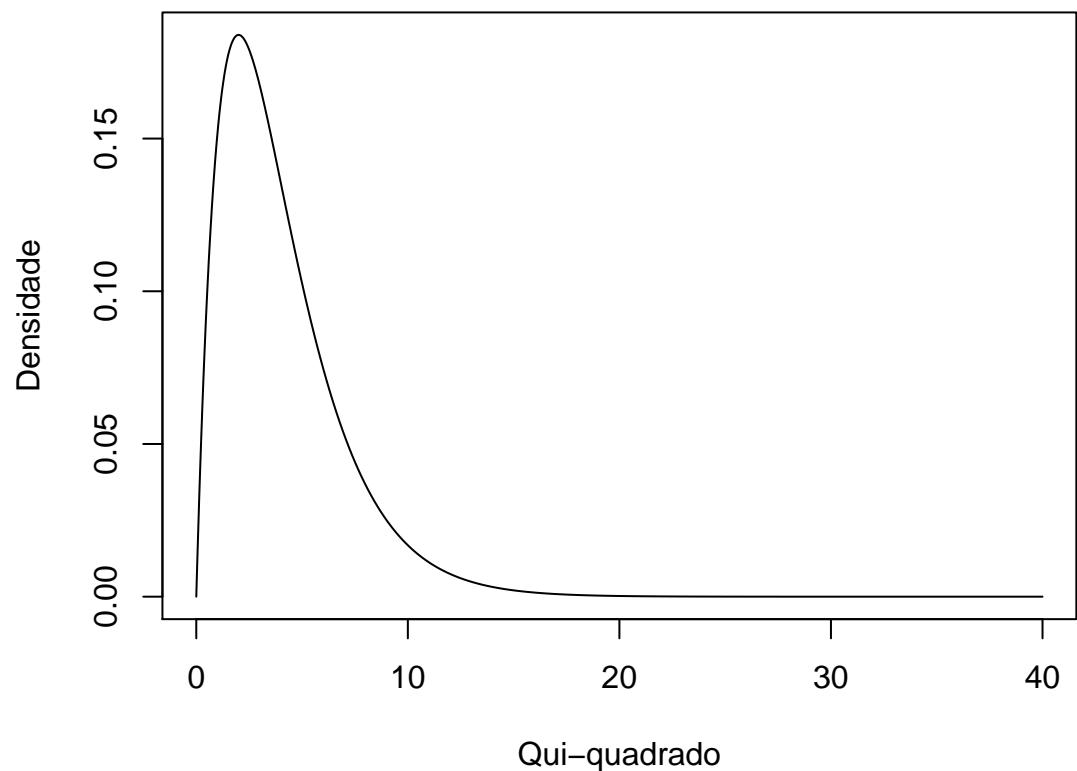
Distribuição de Qui², df=2



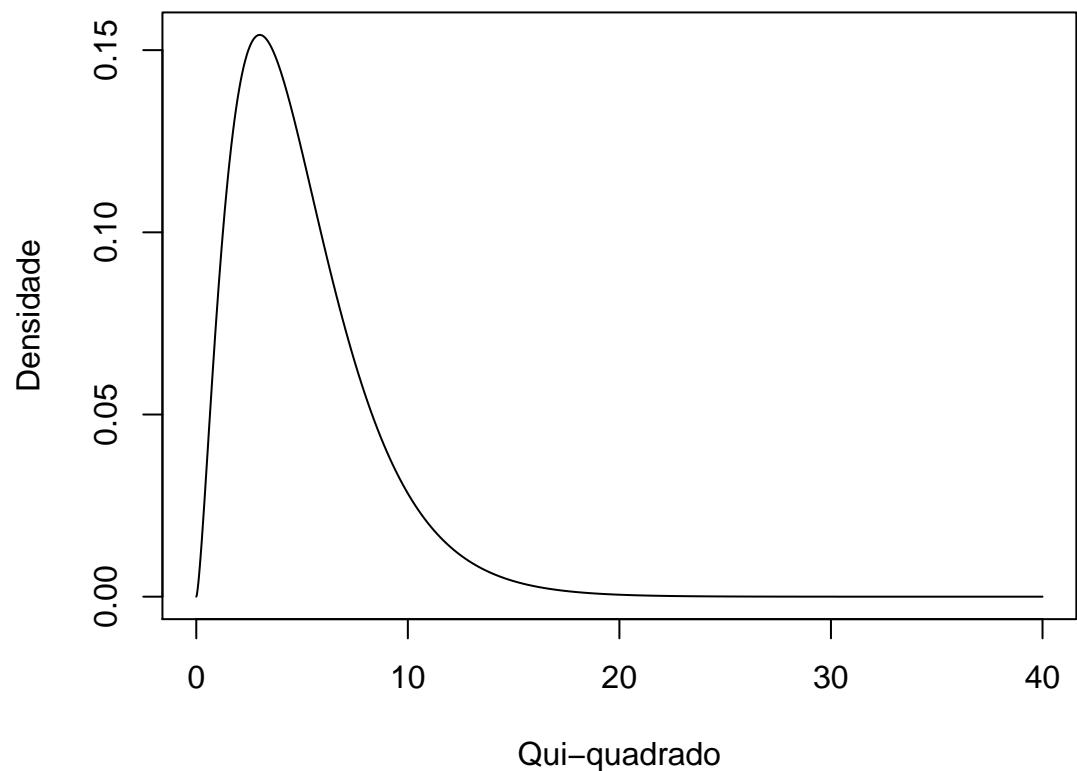
Distribuição de Qui², df=3



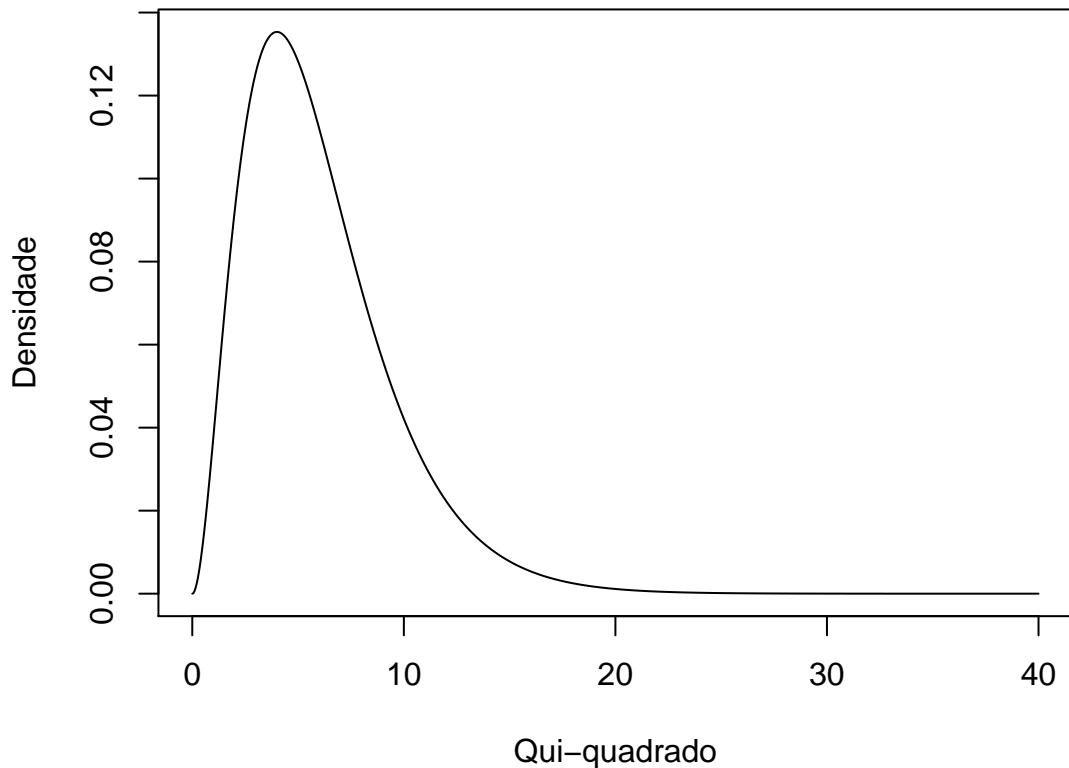
Distribuição de Qui², df=4



Distribuição de Qui², df=5



Distribuição de Qui², df=6



ou, caso prefira ver todas as curvas no mesmo gráfico, sugere-se:

```
source("friendlycolor.R")
cor <- 3
padrao <- 1
leg.txt <- c()
leg.cor <- c()
chi2.valor <- seq(from=0, to=40, by=0.01)
graus.liberdade <- 2:6
for (g in 1:length(graus.liberdade))
{
  p <- dchisq(chi2.valor, df=graus.liberdade[g])
  if (g==1)
  {
    plot(chi2.valor, p, type="l", lty=padrao,
         col=friendlycolor(cor), lwd=2,
         main="Distribuições de Qui^2",
         xlab="Qui-quadrado", ylab="Densidade")

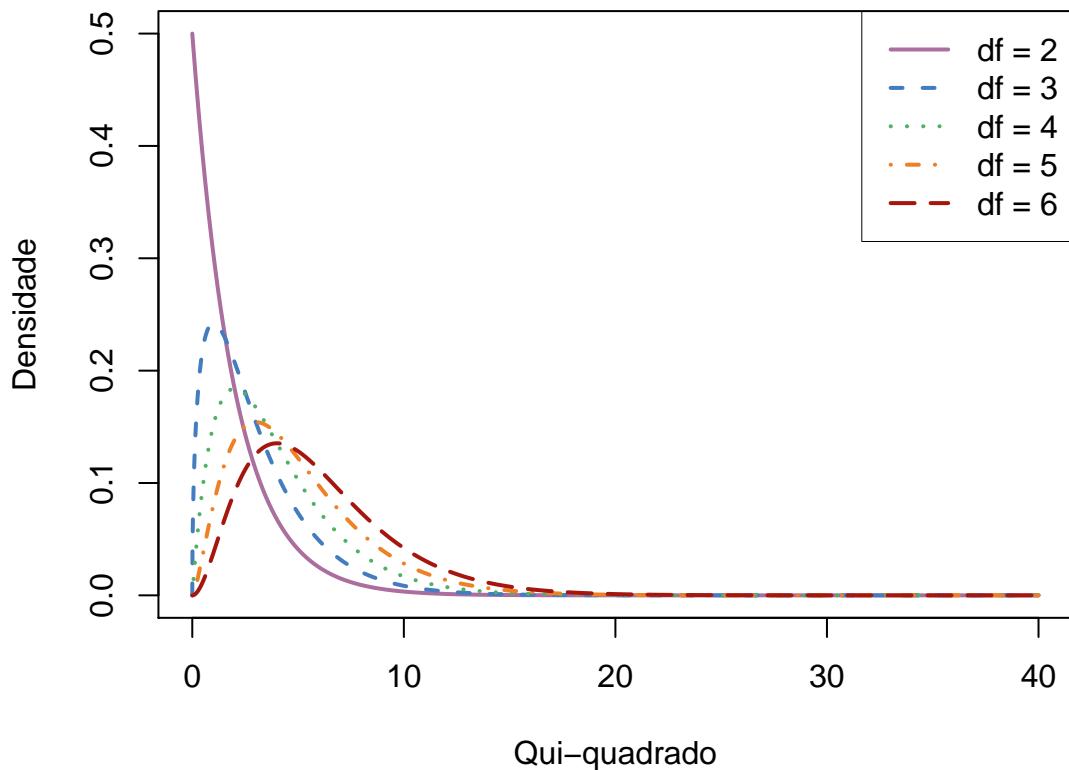
  } else
  {
    lines(chi2.valor, p, lty=padrao,
          col=friendlycolor(cor), lwd=2)
  }
}
```

```

# guarda para a legenda
leg.txt <- c(leg.txt, paste("df = ", graus.liberdade[g], sep=""))
leg.cor <- c(leg.cor, friendlycolor(cor))
cor <- cor+6
padrao <- padrao+1
}
legend("topright",
       leg.txt,
       col=leg.cor,
       lwd=2,
       lty=1:5,
       box.lwd=0, bg="transparent")

```

Distribuições de Qui²



São distribuições assimétricas, iniciadas em zero, que tendem assintoticamente a zero com o valor crescente de χ^2 .

Exemplo: capacete e trauma

Considere um estudo que investiga a efetividade dos capacetes de segurança de bicicleta na prevenção de traumas cranianos. Os dados consistem de uma amostra aleatória de 793 indivíduos envolvidos em acidentes ciclísticos durante um certo período. Deseja-se testar, com $\alpha = 0.05$, se o uso do capacete tem funcionado como um fator de proteção.

Formulamos as hipóteses nula e alternativa:

H_0 : não existe associação entre uso do capacete e trauma craniano

H_1 : existe associação entre uso do capacete e trauma craniano

$\alpha = 5\%$

Os dados são:



```
tbla <- as.table(matrix(c(17, 138, 130, 508), nrow = 2, byrow = TRUE))
colnames(tbla) <- c("Trauma +", "Trauma -")
rownames(tbla) <- c("Capacete +", "Capacete -")
print (tbla)
```

	Trauma +	Trauma -
Capacete +	17	138
Capacete -	130	508

Sob a hipótese nula (H_0), i.e., se não existe associação entre o uso de capacete e o trauma craniano nos acidentes, os valores esperados são:

```
chi2 <- chisq.test(tbla)
print(round(chi2$expected,3))
```

	Trauma +	Trauma -
Capacete +	28.733	126.267
Capacete -	118.267	519.733



Conceptualmente, o teste do χ^2 é um teste de proporções.

É possível transformar em proporções, obtendo:

```
tbla <- as.table(matrix(c(17, 138, 130, 508), nrow = 2, byrow = TRUE))
colnames(tbla) <- c("Trauma +", "Trauma -")
rownames(tbla) <- c("Capacete +", "Capacete -")
tbla_prop <- tabela/sum(tabela)
print (round(tbla_prop,3))
```

	Trauma +	Trauma -
Capacete +	0.021	0.174
Capacete -	0.164	0.641

e, sob H_0 , os valores esperados são:

```
soma_col1 <- sum(tbla_prop[,1])
soma_col2 <- sum(tbla_prop[,2])
soma_lin1 <- sum(tbla_prop[1,])
soma_lin2 <- sum(tbla_prop[2,])
tbla_prosp <- tabela_prop
```

```

tabela_propesp[1,1] <- soma_lin1*soma_col1
tabela_propesp[1,2] <- soma_lin1*soma_col2
tabela_propesp[2,1] <- soma_lin2*soma_col1
tabela_propesp[2,2] <- soma_lin2*soma_col2
print(round(tabela_propesp,3))

```

```

Trauma + Trauma -
Capacete +    0.036    0.159
Capacete -    0.149    0.655

```

Note que o teste em R é feito diretamente com os valores observados (pois o tamanho da amostra importa) e seus resultados foram armazenados, acima, na variável **chi2**. O tipo desta variável é dado por:

```
str(chi2)
```

```

List of 9
$ statistic: Named num 6.7
..- attr(*, "names")= chr "X-squared"
$ parameter: Named int 1
..- attr(*, "names")= chr "df"
$ p.value : num 0.00964
$ method   : chr "Pearson's Chi-squared test with Yates' continuity correction"
$ data.name: chr "tabela"
$ observed : 'table' num [1:2, 1:2] 17 130 138 508
..- attr(*, "dimnames")=List of 2
... .$. : chr [1:2] "Capacete +" "Capacete -"
... .$. : chr [1:2] "Trauma +" "Trauma -"
$ expected : num [1:2, 1:2] 28.7 118.3 126.3 519.7
..- attr(*, "dimnames")=List of 2
... .$. : chr [1:2] "Capacete +" "Capacete -"
... .$. : chr [1:2] "Trauma +" "Trauma -"
$ residuals: 'table' num [1:2, 1:2] -2.189 1.079 1.044 -0.515
..- attr(*, "dimnames")=List of 2
... .$. : chr [1:2] "Capacete +" "Capacete -"
... .$. : chr [1:2] "Trauma +" "Trauma -"
$ stdres   : 'table' num [1:2, 1:2] -2.7 2.7 2.7 -2.7
..- attr(*, "dimnames")=List of 2
... .$. : chr [1:2] "Capacete +" "Capacete -"
... .$. : chr [1:2] "Trauma +" "Trauma -"
- attr(*, "class")= chr "htest"

```

```
sapply(chi2, typeof)
```

statistic	parameter	p.value	method	data.name	observed
"double"	"integer"	"double"	"character"	"character"	"double"
expected	residuals	stdres			
"double"	"double"	"double"			

É uma estrutura complexa, da qual muitos valores podem ser extraídos quando quisermos utilizá-los dentro de um *RScript*. A parte em **chi2\$expected**, que já utilizamos, contém a tabela com os valores esperados sob H_0 .

O resultado principal do teste é exibido pela variável como um todo:

```
print (chi2)
```

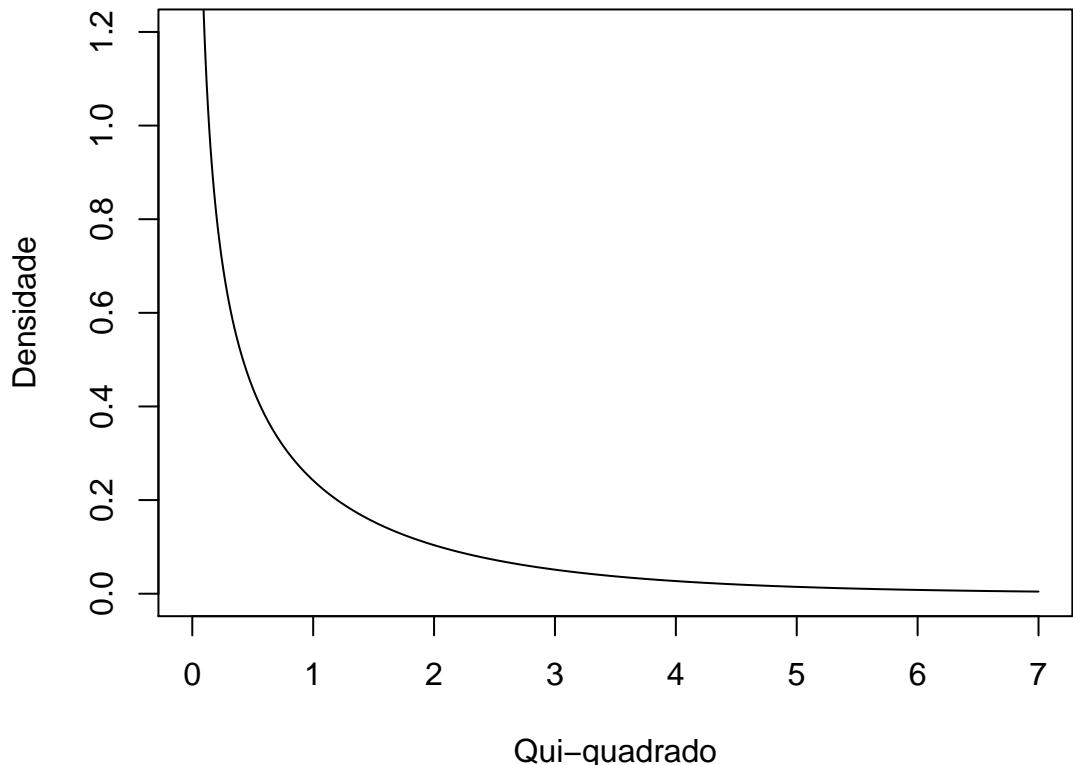
Pearson's Chi-squared test with Yates' continuity correction

```
data: tabela  
X-squared = 6.7001, df = 1, p-value = 0.009641
```

A estatística de teste tem o valor $\chi^2 = 6.7001$, com 1 grau de liberdade ($df = 1$), correspondendo ao valor-p de 0.009641 (que será discutido adiante). Com estes valores podemos decidir pela rejeição ou não-rejeição de H_0 .

No exemplo de uma tabela 2x2 há somente 1 grau de liberdade. O gráfico da distribuição de qui-quadrado (para melhor visibilidade, computamos $0 \leq \chi^2 \leq 7$ e truncamos o eixo y em 1.2) é:

```
chi2.valor <- seq(from=0, to=7, by=0.01)  
graus.liberdade <- 1  
p <- dchisq(chi2.valor, df=graus.liberdade)  
plot(chi2.valor, p, type="l", ylim=c(0, 1.2),  
     xlab="Qui-quadrado", ylab="Densidade"  
)
```



Nosso interesse é saber a partir de qual o valor de χ^2 a discrepância entre os valores observados e esperados é menos provável que o valor de α escolhido. Isto é encontrar o valor que separa a área de 5% sob a curva de distribuição em sua cauda direita. Este valor é chamado de qui-quadrado crítico, χ_c^2 .



Note que o raciocínio é bicaudal, buscando associação positiva ou negativa, mas a operacionalização do teste só utiliza a cauda superior. Isto acontece porque $\chi^2 = 0$ indica igualdade numérica (e, portanto, também estatística) entre os valores esperados e observados. Consequentemente, procuramos $\alpha = 0.05$ somente na cauda direita.

Em R, podemos encontrar χ_c^2 pelo complemento ($1 - \alpha = 0.95$) com

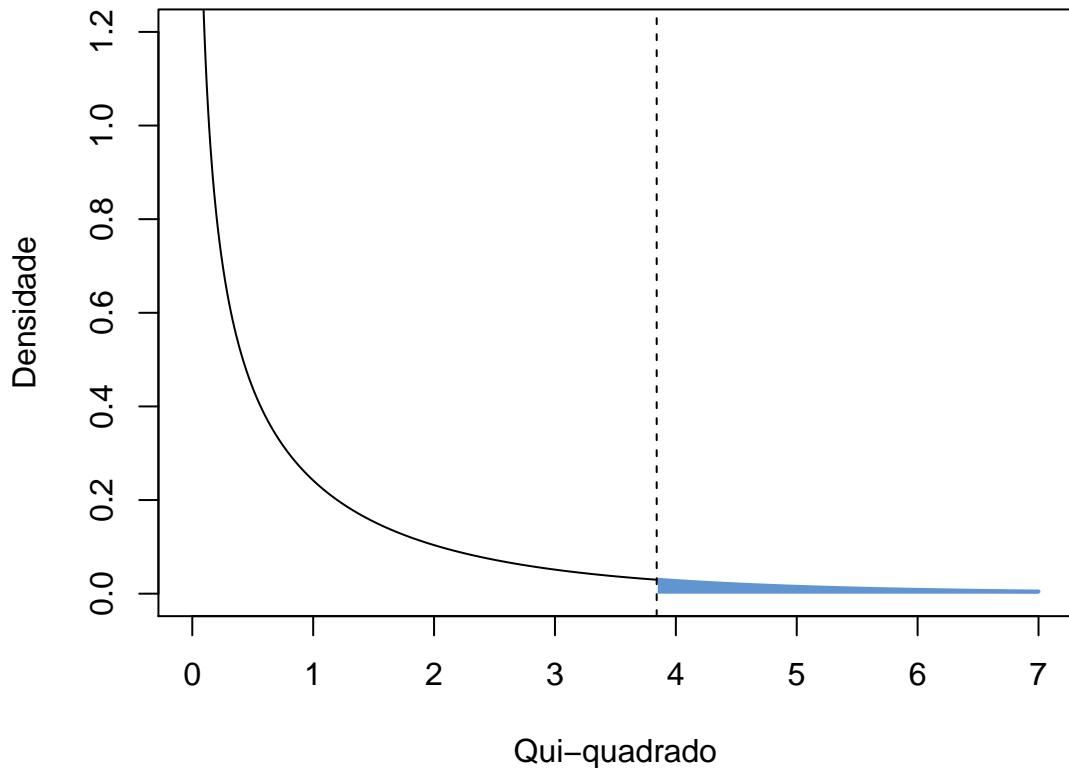
```
chi2.critico <- qchisq(p=0.95, df=1)
cat("qui-quadrado crítico (df=1, alfa=5%): ", chi2.critico, "\n", sep="")
```

qui-quadrado crítico (df=1, alfa=5%): 3.841459

pois este é o valor de χ^2 que deixa 95% da área sob a curva à sua esquerda.

Graficamente, podemos hachurar em azul a área correspondente a α e traçar uma linha pontilhada vertical em χ_c^2 com:

```
source("friendlycolor.R")
# vetor com valores de 0 a 7
chi2.valor <- seq(from=0, to=7, by=0.01)
graus.liberdade <- 1
# vetor com probabilidades correspondentes
p <- dchisq(chi2.valor, df=graus.liberdade)
# exibe o gráfico
plot(chi2.valor, p, type="l", ylim=c(0, 1.2),
      xlab="Qui-quadrado", ylab="Densidade")
# vetor com os valores acima de qui critico
chi2.critico <- qchisq(p=0.95, df=1)
chi2maioralfa <- chi2.valor[chi2.valor > chi2.critico]
# probabilidades correspondentes
pmaioralfa <- dchisq(chi2maioralfa, df=graus.liberdade)
# hachura sob a curva
chi2maioralfa <- c(min(chi2maioralfa), chi2maioralfa, max(chi2maioralfa))
pmaioralfa <- c(0, pmaioralfa, 0)
polygon(chi2maioralfa, pmaioralfa, col=friendlycolor(10), border = NA)
# reforça a linha
lines(chi2maioralfa[2:(length(chi2maioralfa)-1)],
      pmaioralfa[2:(length(pmaioralfa)-1)],
      col=friendlycolor(10), lwd=2)
# linha pontilhada, valor critico
abline(v=chi2.critico, lty=2)
```



Esta área hachurada é a região de rejeição de H_0 .

- decisão pela estatística de teste

A estatística χ^2 é maior quanto maior for a discrepância entre os valores observados e esperados. Para um grau de liberdade e $\alpha = 0.05$, computamos $\chi_c^2 = 3.8414588$. No exemplo de capacete e trauma, acima, rejeitamos H_0 porque calculamos $\chi^2 = 6.7001206$, valor maior que o valor crítico.

- decisão pelo valor-p

À direita do valor calculado $\chi^2 = 6.7001206$ define-se uma área, que corresponde ao valor-p. Este valor também foi calculado quando aplicamos o teste e guardamos seu resultado em `chi2`, acima; o valor-p encontra-se acessível em `chi2$p.value=0.0096406`. Como $p < \alpha$, concluimos pela rejeição de H_0 .



As duas formas de decisão são igualmente válidas e equivalentes.

Exibir o valor-p é a forma mais moderna e recomendada atualmente.

Teste do χ^2 robusto

Repare nas saídas da função `chisq.test()`, cujo cabeçalho indica:

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

Esta é uma versão tradicional do teste, para a qual a correção de Yates é aplicável por *default* às tabelas 2x2 (mais detalhes serão vistos adiante). Podemos evitá-la com:

```
chi2ny <- chisq.test(tabela, correct=FALSE)
print(chi2ny)
```

Pearson's Chi-squared test

```
data: tabela
X-squared = 7.3099, df = 1, p-value = 0.006858
```

Além disto, há várias condições para a aplicação do teste do qui-quadrado (comentadas adiante) que podem, muitas vezes, restringir seu uso.

Uma alternativa é computar o teste em sua versão robusta. A função `chisq.test()` consegue simular a distribuição por *bootstrapping*, alterando-se o comando para:

```
chi2.robusto <- chisq.test(tabela, correct=FALSE,
                             simulate.p.value = TRUE, B=1e6)
print(chi2.robusto)
```

Pearson's Chi-squared test with simulated p-value (based on 1e+06
replicates)

```
data: tabela
X-squared = 7.3099, df = NA, p-value = 0.007744
```

Esta versão é simulada com $1e6 = 1 \cdot 10^6 = 1000000$ iterações.



Como a amostra é grande, os valores de χ^2 e p não mudam muito entre as diferentes versões. Há um pequeno *bug* na implementação da versão robusta, que não calcula os graus de liberdade (sabemos que $df = 1$).

A decisão, neste caso, não mudou, tanto pelo critério do χ^2_c quanto pelo p : rejeita-se H_0 .

Riscos relativos



Odds e probabilidade

Odds (traduzido habitualmente por chance) e probabilidade são formas equivalentes para expressar possibilidades e relacionadas por:

$$Odds = \frac{\text{Probabilidade}}{1 - \text{Probabilidade}}$$

e

$$\text{Probabilidade} = \frac{Odds}{1 + Odds}$$

Por exemplo, uma probabilidade de 80% corresponde a

$$Odds = \frac{0.8}{1 - 0.8} = \frac{0.8}{0.2} = 4$$

indicando que dizer “80% de probabilidade” é equivalente a dizer “4 vezes mais chance de ocorrer do que não ocorrer” um determinado evento.

Resersamente, *Odds* de 2 é:

$$\text{Probabilidade} = \frac{2}{1+2} = \frac{2}{3} \approx 66.67\%$$

e *Odds* = 1 resulta em:

$$\text{Probabilidade} = \frac{1}{1+1} = \frac{1}{2} = 50\%$$

Associa-se o máximo de incerteza com probabilidade de 50%; por este motivo, *Odds* = 1 será um valor necessário para decisões estatísticas, adiante.

Há dois tipos principais de risco relativo: *razão de riscos* (RR, *risk ratio*) e *razão de chances* (OR, *odds ratio*).

Vamos considerar a seguinte tabela:

	Tem efeito	Não tem efeito	
Tem exposição	a	b	a+b
Não tem exposição	c	d	c+d
	a+c	b+d	total

Esta tabela de contingência genericamente relaciona exposição (e.g., hábito de fumar) com o efeito (e.g., câncer de pulmão). Traz as contagens dos:

- expostos que apresentaram o efeito (*a*),
- expostos que não apresentaram o efeito (*b*),
- não expostos que apresentaram o efeito (*c*) e
- não expostos que não apresentaram o efeito (*d*),

razão de riscos (RR, risk ratio)

Riscos são probabilidades. A razão de riscos é dada por:

$$RR = \frac{\frac{a}{a+b}}{\frac{c}{c+d}}$$

Repare que $\frac{a}{a+b}$ é a probabilidade de um indivíduo exposto ter o efeito (e.g., fumante ter câncer); $\frac{c}{c+d}$ é a probabilidade de um indivíduo **não** exposto ter o efeito (e.g., não-fumante ter câncer).

RR, portanto, é um *odds*: quantas vezes mais é esperado o efeito (ocorrência de câncer) de quem é exposto (fumante) em comparação com quem não é exposto (não fumante);

razão de chances (OR, odds ratio)

Odds ratio como diz o nome, é uma razão entre dois *odds*, dada por:

$$OR = \frac{\frac{a}{a+c}}{\frac{b}{b+d}}$$

Considere, agora, que:

- $\frac{a}{a+c}$, entre os doentes, quantas vezes é a maior a chance (*odds*) de ter efeito (câncer) para quem é exposto (fuma).
- $\frac{b}{b+d}$, entre os não doentes, quantas vezes é a maior a chance (*odds*) de ter efeito (câncer) para quem é exposto (fuma).

Interessante, também, é que (com alguma álgebra) *OR* reduz-se a:

$$OR = \frac{a \cdot d}{b \cdot c}$$

fácil de lembrar e assim conhecido como “produto cruzado” (com o defeito de ocultar o motivo pelo qual é um *odds ratio*). Por outro lado, como produto cruzado, percebe-se que é uma medida de quanto pesa a:

- diagonal principal, do que é concordante: efeito entre expostos E não efeito entre não expostos em relação à
- diagonal secundária, do que é discordante: efeito entre não expostos E não efeito entre expostos.



Os valores de *RR* e *OR* são próximos quando menos que 5% dos não expostos têm efeito.

Em R, o cálculo de *OR* aparece no Teste Exato de Fisher. No exemplo de capacete e trauma, temos:

```
tbla <- as.table(matrix(c(17, 138, 130, 508), nrow = 2, byrow = TRUE))
colnames(tbla) <- c("Trauma +", "Trauma -")
rownames(tbla) <- c("Capacete +", "Capacete -")
print(tbla)
ft <- fisher.test(tbla) # Teste de OR robusto
print(ft)
```

	Trauma +	Trauma -
Capacete +	17	138
Capacete -	130	508

Fisher's Exact Test for Count Data

```
data: tbla
p-value = 0.005688
alternative hypothesis: true odds ratio is not equal to 1
95 percent confidence interval:
0.2630993 0.8357342
sample estimates:
odds ratio
```

0.4817645

- decisão pelo intervalo de confiança

A estimativa pontual de *odds ratio* está em `ft$estimate=0.4817645`, mas para a decisão é obrigatório verificar seu intervalo de confiança 95% (*IC95*), com limites inferior e superior respectivamente guardados em `ft$conf.int[1]=0.2630993` e `ft$conf.int[2]=0.8357342`.

Como a ausência de efeito é dada por $OR = 1$ (correspondendo a H_0) e este valor unitário está fora do intervalo, rejeita-se H_0 . Além disto, o *IC95* é menor que o esperado por H_0 e, portanto, o uso de capacete está associado com redução dos traumas cranianos.



Caso a tabela fosse construída com as colunas em posições trocadas, obter-se-ia o seguinte:

	Trauma -	Trauma +
Capacete +	138	17
Capacete -	508	130

Fisher's Exact Test for Count Data

```
data: tabela
p-value = 0.005688
alternative hypothesis: true odds ratio is not equal to 1
95 percent confidence interval:
1.196553 3.800847
sample estimates:
odds ratio
2.075703
```

O *odds ratio* é agora maior que o valor unitário, o qual está fora do intervalo, rejeitando-se a hipótese nula e a associação é positiva: a conclusão é a mesma, o uso de capacete está associado com o aumento de “Trauma -”, portanto é protetor.

Aliás, observe que

```
print(1/ft$estimate)

odds ratio
0.4817645
```

```
print(1/ft$conf.int)
```

```
[1] 0.8357342 0.2630993
attr(",conf.level")
[1] 0.95
```

são os mesmos valores obtidos anteriormente.

Dois códigos úteis, com vários dos elementos apresentados neste texto

4 Simulação e Gráficos

tabelacontingencia2x2_capacetetrauma.R

Em *tabelacontingencia2x2_capacetetrauma.R*, que você pode adaptar para as suas necessidades, processamos o mesmo exemplo (procure a função *sink()* que desvia o resultado da tela para um arquivo texto). Além da estatística χ^2 , computamos outros valores interessantes. Observe sua saída (*tabelacontingencia2x2_capacetetrauma.txt*):

```
source("tabelacontingencia2x2_capacetetrauma.R")
```

Resultados armazenados em *tabelacontingencia2x2_capacetetrauma.txt*

Em *Qui2.R* também processamos o mesmo exemplo. Esta versão aplica bootstrapping de um jeito mais manual e produz gráficos associados aos conceitos que discutimos.

Execute com:

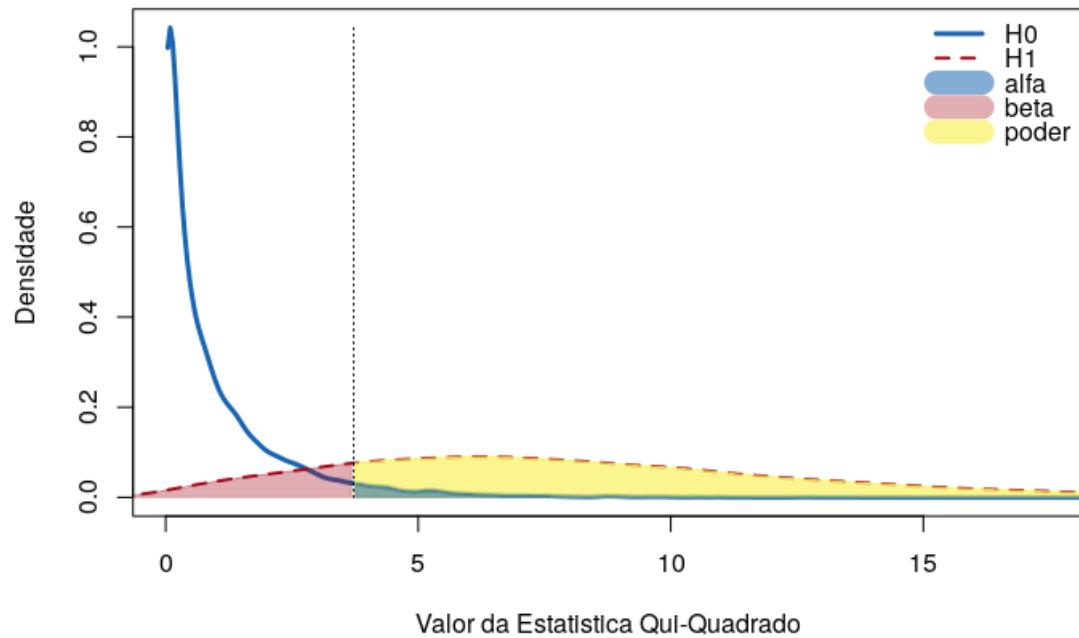
```
source("Qui2.R")
```

fornecendo:

Observe a simulação em andamento na área de *Plots*. Quando terminar, procure os gráficos finalizados. Os principais são:

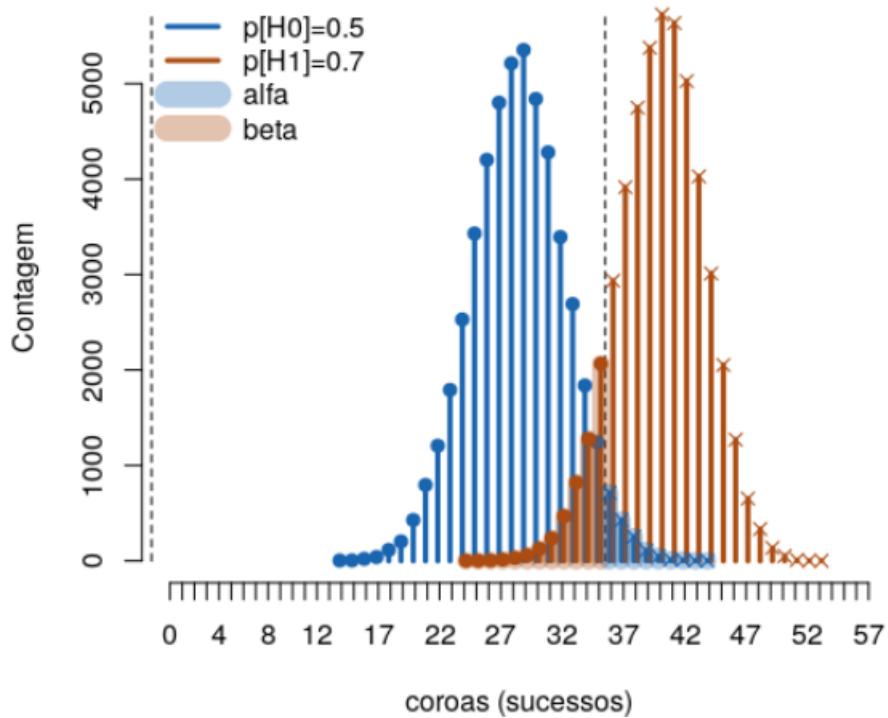
As distribuições de χ^2 para a hipótese nula e para o valor observado de H_1 (o valor de β e o poder associados, aqui, são os valores *a posteriori*, portanto inúteis para qualquer decisão):

Capacete x Trauma (n=793, 10000 iteracoes)



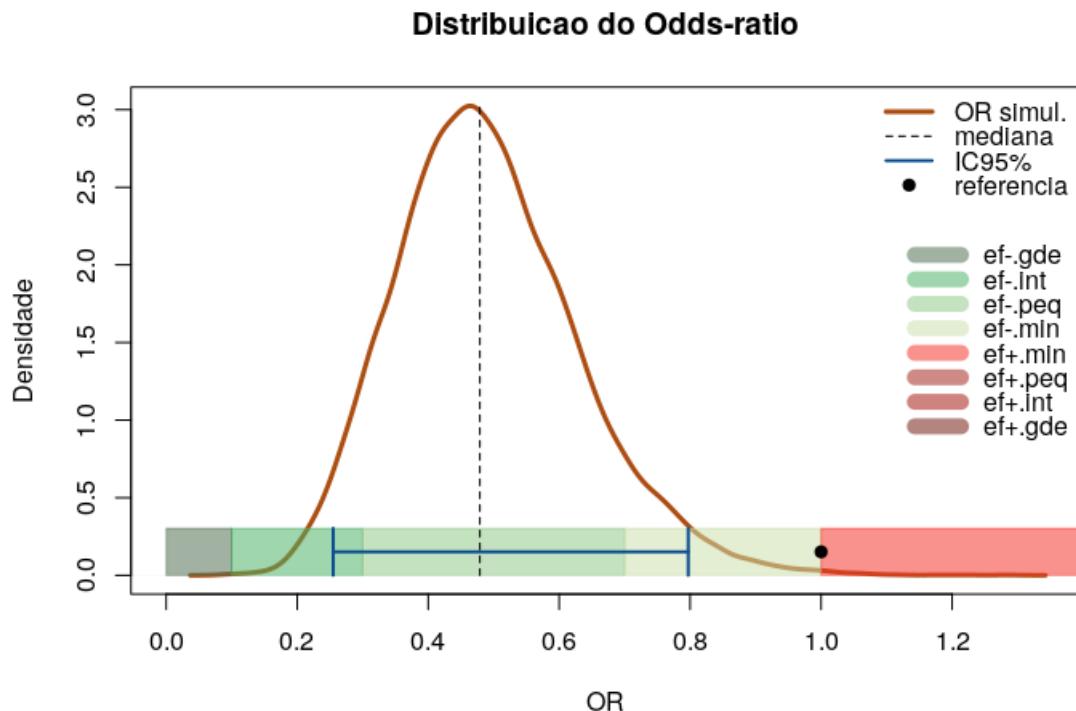
Flashback:

Recorde a tomada de decisão com a distribuição binomial:



A estatística e o formato das distribuições mudam (em função, também, do tipo de variável), mas as áreas de α , β e o raciocínio são os mesmos.

A distribuição, com intervalo de confiança, do *odds ratio* simulado (compare com o obtido acima), indicando a rejeição de H_0 (o valor 1, de referência, está fora do intervalo) e que seu efeito está entre mínimo a intermediário:



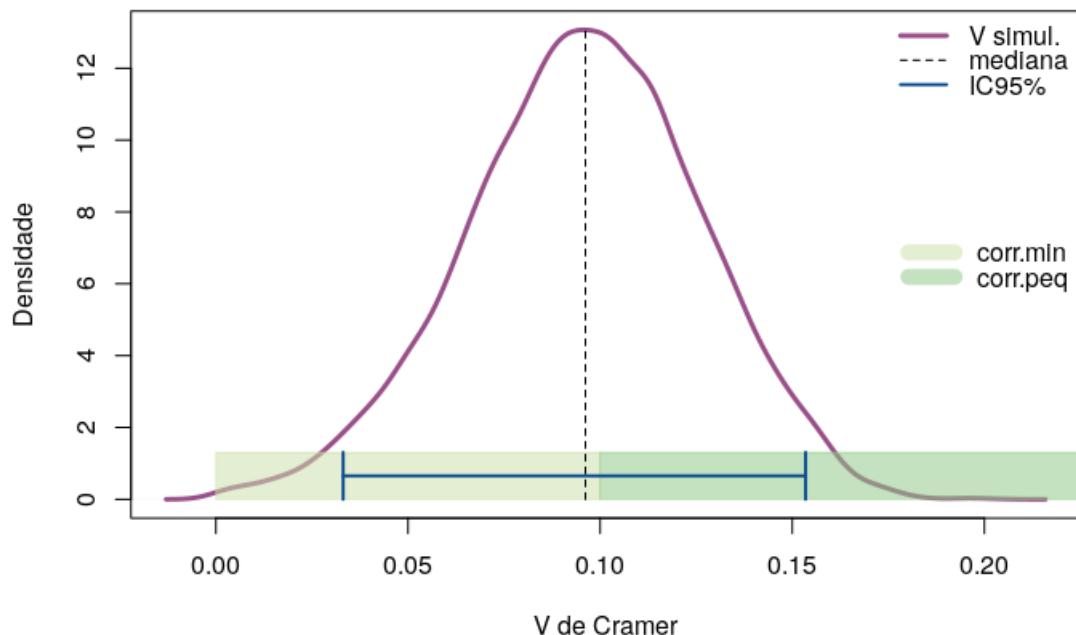
Odds ratio é uma medida da intensidade de associação entre duas variáveis nominais dicotômicas de exposição e desfecho.

De acordo com SHARPE, D. (2015) Your chi-square test is statistically significant: now what? Practical Assessment, Research & Evaluation, 20(8). É, também, uma medida de tamanho de efeito (não depende do tamanho da amostra, é adimensional e adirecional) e sua intensidade pode ser classificada em:

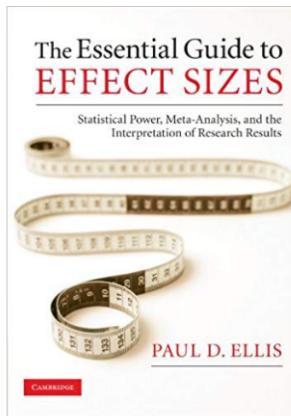
- OR para categoria de proteção
 - Grande: $OR < 0,1$
 - Intermediária: $0,1 \leq OR < 0,3$
 - Pequena: $0,3 \leq OR < 0,7$
 - Mínima: $0,7 \leq OR < 1$
- OR para categoria de risco
 - Mínima: $1 < OR \leq 1,5$
 - Pequena: $1,5 < OR \leq 3,5$
 - Intermediária: $3,5 < OR \leq 9$
 - Grande: $OR > 9$

A distribuição do tamanho de efeito (V de Cramer), mostrando que o efeito, apesar de significante, está entre mínimo e pequeno:

Distribuição do V de Cramer



O programa já indica o tamanho de efeito de acordo com



Capa comum: 173 páginas
Editora: Cambridge University Press; Edição: 1 (1 de setembro de 2010)
Idioma: Inglês
ISBN-10: 0521142466
ISBN-13: 978-0521142465

O V de Cramer é uma medida do grau de correlação absoluta entre duas variáveis nominais:

- não depende do total de observações independentes
- não depende do número de linhas e colunas da TC

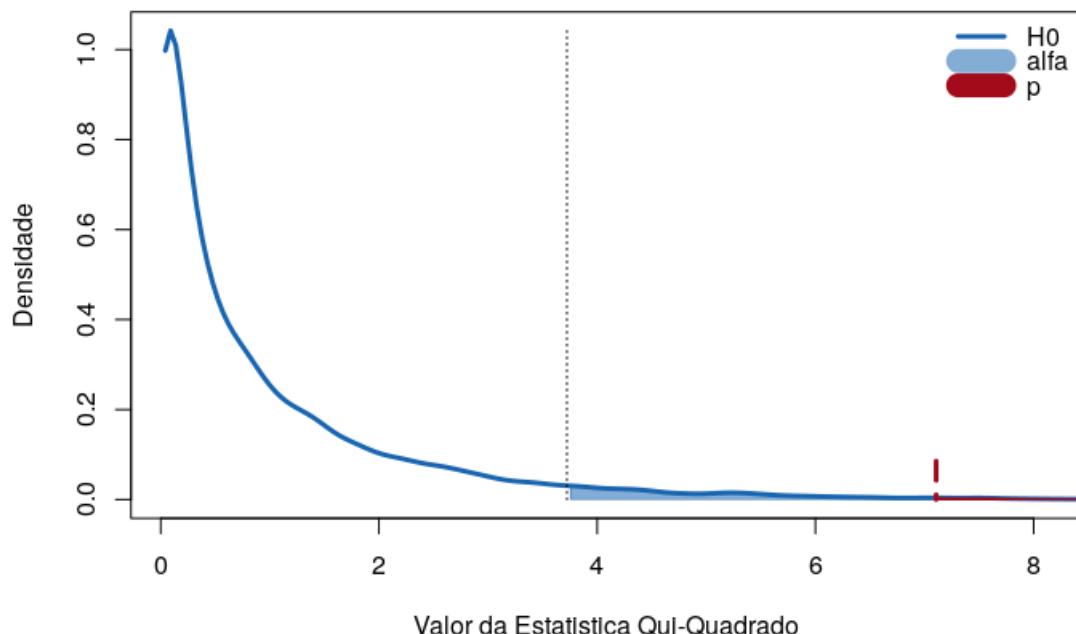
- é adimensional e varia entre 0 (independência) e 1 (dependência)

Tamanho de efeito:

- $0 \leq V < 0.1$: mínimo
- $0.1 \leq V < 0.3$: pequeno
- $0.3 \leq V < 0.5$: intermediário
- $V \geq 0.5$: grande

Novamente a distribuição de χ^2 , mas aqui exibe somente H_0 e o valor-p, indicando a rejeição de H_0 :

Capacete x Trauma (n=793, 10000 Iteracoes)



Tabelas LxC

O teste do χ^2 opera em tabelas maiores, com L linhas e C colunas.

Por exemplo, três quimioterápicos foram comparados quanto ao efeito colateral (náusea). Considere $\alpha = 0.05$. Há diferença entre as drogas?

Os dados obtidos foram:

Em *Qui2_LxC.R* implementamos este exemplo (versão robusta):

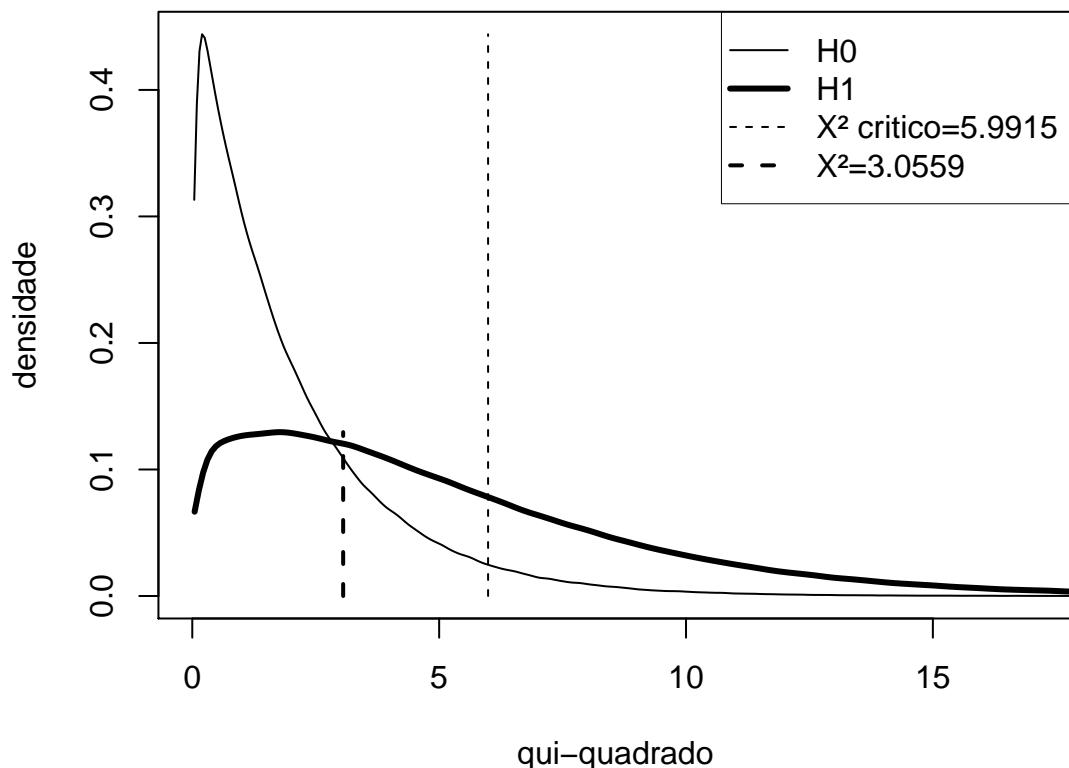
```
source("Qui2_LxC.R")
```

	ComNausea	SemNausea
A	3	5
B	7	2
C	6	3

Pearson's Chi-squared test with simulated p-value (based on 1e+06 replicates)

```
data: tcrxc
X-squared = 3.0559, df = NA, p-value = 0.2739
```

```
V de Cramer = 0.3428334
```



Observe que não rejeita-se H_0 . Os dois critérios são sempre equivalentes:

- o χ^2 calculado é menor que χ_c^2
- $p > \alpha$

Abra o código para ver como foi implementado este teste.

Métodos tradicionais

5 Condições para o χ^2 (métodos não robustos)

Antes das implementações computacionais e a facilidade do uso do R, a partir de uma tabela 2x2 era necessário construir a tabela dos valores esperados.

- Valores observados:

- Valores esperados:

Cada valor esperado é obtido pelo produto das marginais, dividido pelo total geral:

$$\begin{aligned} - (147 \cdot 155) / 793 &\approx 28.7 \\ - (147 \cdot 638) / 793 &\approx 118.3 \\ - (646 \cdot 155) / 793 &\approx 126.3 \\ - (646 \cdot 638) / 793 &\approx 519.7 \end{aligned}$$

Avaliando os dados observados e os valores esperados, era necessário verificar se o teste de χ^2 podia ser aplicado. Eram as seguintes condições:

SIEGEL, S. & CASTELLAN Jr., N.J. (1988) Nonparametric statistics for the behavioral sciences. 2^a ed. NY: McGraw-Hill, p. 123:

- O teste exato de Fisher para testar independência é adequado apenas para tabela de contingência 2×2 se $N < 20$
- Se N entre 20 e 40, o teste qui-quadrado de Pearson pode ser usado se todas as frequências esperadas são maiores que 5.
- Se $N > 40$, o teste qui-quadrado de Pearson com correção de Yates pode ser usado.

COCHAN, W.G. (1954) Some Methods for Strengthening the Common χ^2 Tests. Biometrics: 10(4): 417-451:

- Um valor esperado mínimo de 1 em alguma célula é permitido, desde que não mais que 20% das células tenham valor abaixo de 5

Então, o valor de χ^2 era dado por

$$\chi^2 = \sum \frac{(\text{observado} - \text{esperado})^2}{\text{observado}}$$

neste exemplo calculado por:

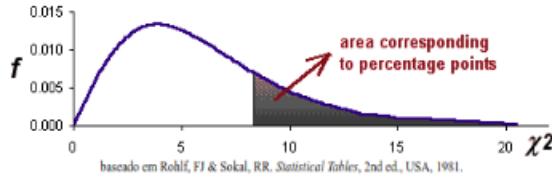
$$\chi^2 = \frac{(17 - 28.7)^2}{28.7} + \frac{(138 - 126.3)^2}{126.3} + \frac{(130 - 118.3)^2}{118.3} + \frac{(508 - 519.7)^2}{519.7} \approx 7.31$$

ou, em tabelas 2x2, era recomendado usar a correção de continuidade de Yates, dada por:

$$\chi^2 = \sum \frac{(|\text{observado} - \text{esperado}| - 0.5n)^2}{\text{observado}}$$

que, neste exemplo, resulta em $\chi^2 \approx 6.7$

Então, era necessário consultar uma tabela com os valores críticos de χ^2_c previamente calculados para a tomada de decisão (sem obter, portanto, o valor- p). A tabela permitia escolher o valor de α e, sabendo-se os graus de liberdade (neste caso $\nu = 1$), localizar χ^2_c (neste exemplo igual a 3.841):



α	0.995	0.975	0.9	0.5	0.1	0.05	0.025	0.01	0.005	0.001	
V	1	0.000	0.001	0.016	0.455	2.706	3.841	5.024	6.635	7.879	10.827
	2	0.010	0.051	0.211	1.386	4.605	5.991	7.378	9.210	10.597	13.815
	3	0.072	0.216	0.584	2.366	6.251	7.815	9.348	11.345	12.838	16.266
	4	0.207	0.484	1.064	3.357	7.779	9.488	11.143	13.277	14.860	18.466
	5	0.412	0.831	1.610	4.351	9.236	11.070	12.832	15.086	16.750	20.515
	6	0.676	1.237	2.204	5.348	10.645	12.592	14.449	16.812	18.548	22.457
...											
114	78.862	86.342	95.128	113.334	133.729	139.921	145.441	152.037	156.637	166.406	
115	79.691	87.213	96.043	114.334	134.813	141.030	146.571	153.190	157.808	167.609	
116	80.522	88.084	96.958	115.334	135.898	142.138	147.700	154.344	158.977	168.813	
117	81.353	88.955	97.874	116.334	136.982	143.246	148.829	155.497	160.146	170.014	
118	82.185	89.827	98.790	117.334	138.066	144.354	149.957	156.648	161.314	171.216	
119	83.018	90.700	99.707	118.334	139.149	145.461	151.084	157.799	162.481	172.417	
120	83.852	91.573	100.624	119.334	140.233	146.567	152.211	158.950	163.648	173.618	

Caso χ^2 não atendesse as exigências para ser aplicado, a alternativa (se fosse tabela 2x2) era o Teste Exato de Fisher. Este envolve calcular o valor- p por fatoriais de tabelas progressivamente mais extremas. Por exemplo:

Observado:		Possível efeito			
		Sim	Não		
Exposição	Sim	3	9	12	
	Não	7	1	8	
		10	10	20	

Esperado:		Possível efeito			
		Sim	Não		
Exposição	Sim	6	6	12	
	Não	4	4	8	
		10	10	20	

Não pode Usar χ^2

A probabilidade de obtermos esta tabela de frequência observada é dada por:

$$\begin{array}{cc} 3 & 9 \\ 7 & 1 \end{array} \quad p_1 = \frac{(a+b)!(c+d)!(a+c)!(b+d)!}{N!a!b!c!d!} = \frac{12!8!10!10!}{20!3!9!7!1!} = 0,00953$$

Reduzir de 1 a casela com menor frequência:

$$\begin{array}{cc} 3 & 9 \\ 7 & 1 \end{array} \rightarrow \begin{array}{cc} 2 & 10 \\ 8 & 0 \end{array} \quad p_2 = \frac{12!8!10!10!}{20!2!10!8!0!} = 0,000357$$

$$p' = p_1 + p_2 = 0,0095261 + 0,000357 = 0,0098838$$

$$p = 2 \times 0,0098838 = 0,019766$$

Neste caso o valor exato de p era calculado e comparado diretamente com α para a tomada de decisão.

O que diz a literatura mais recente sobre estes métodos tradicionais e não robustos?

Is there still a place for Pearson's chi-squared test and Fisher's exact test in surgical research?

© 2011 The Author

ANZ Journal of Surgery © 2011 Royal Australasian College of Surgeons

ANZ J Surg 81 (2011)

Table 2A Number of articles containing the listed phrases published in the decade 2001–2011 in three surgical journals, according to Google Scholar, together with Impact Factors

Phrase	ANZ J. Surg. (IF: 1.34)	Surgery (IF: 3.60)	Journal Br. J. Surg. (IF: 4.08)	Ann. Surg. (IF: 7.90)
Chi-squared test	109	1700	371	141
Fisher's exact test	66	8580	462	1490
Fisher–Freeman–Halton	1	20	2	5
Cochran–Armitage	2	149	6	31

IF, Impact Factors 2010.

Conclusions

I cannot do other than conclude that Pearson's chi-squared test and Fisher's exact test should almost never be used to analyse the results of surgical studies. The grounds for this conclusion are as follows:

- (1) These tests have prerequisites that can rarely, if ever, be fulfilled in real-life studies (see Table 3).
- (2) Both these procedures test very non-specific hypotheses (see Table 3). It is much better to use procedures that are designed to detect inequality of proportions, a ratio of proportions that differs from unity or a trend in proportions.
- (3) It is better to use exact permutation tests rather than approximate tests based on the chi-squared or normal distributions.

