Análise de duas variáveis qualitativas

Medidas de Associação

- Razão de prevalências (estudos de prevalência ou transversais)
- Razão de incidências ou risco relativo (estudos de coorte ou de seguimento)
- Razão de odds ou odds ratio (estudos caso-controle)
- Qui quadrado de Pearson e coeficiente de associação de Yule (estudos de prevalência, incidência e caso-controle).

Razão de prevalências

Estudo de prevalência

São apresentados dados sobre o estado nutricional de 1226 crianças brasileiras de 2 anos de idade, segundo sexo. Local X, Ano Y.

Estado nutricional	Masculino	Feminino	Total
Desnutridas	29	20	49
Normais	574	603	1177
Total	603	623	1226

Fonte: dados hipotéticos.

Prevalência de desnutrição: $\frac{49}{1226} = 0{,}040$ ou 4%.

Prevalência de desnutrição segundo sexo:

Masculino: $\frac{29}{603} = 0{,}05$ ou 5,0%; Feminino: $\frac{20}{623} = 0{,}032$ ou 3,2%.

Razão de prevalências: $\frac{\frac{29}{603}}{\frac{20}{623}} = 1{,}5$

Se a razão de prevalências for igual a 1 ou a diferenças de prevalências for igual a 0 então se diz que as variáveis não estão associadas. Na inferência estatística é possível testar se o valor observado vem de uma população com parâmetro igual a 1. Até agora estamos construindo a estatística que permite verificar associação.

Interpretação como medida de efeito:

A prevalência de desnutrição entre meninos é 1,5 vezes (uma vez e meia) a prevalência de desnutrição entre meninas.

Razão de prevalências: $\dfrac{\frac{29}{603}}{\frac{20}{623}} = 1{,}5$

A prevalência de desnutrição parece ser maior entre as crianças do sexo masculino. Os meninos apresentam uma prevalência 50% maior do que as meninas, calculado como (1,5-1)x100.

Se para o cálculo da razão de prevalências for utilizada a prevalência entre meninas no numerador, a razão de prevalências se altera para:

Razão de prevalências: $\dfrac{\frac{20}{623}}{\frac{29}{603}} = \dfrac{603x20}{623x29} = 0{,}68$

De modo semelhante,

Se a razão de prevalências for igual a 1 então se diz que as variáveis não estão associadas. Na inferência estatística é possível testar se o valor observado vem de uma população com parâmetro igual a 1.

Interpretação como medida de efeito:

A prevalência de desnutrição entre meninas é 0,68 vezes a prevalência de desnutrição entre meninos.

A prevalência de desnutrição parece ser menor entre as crianças do sexo feminino. As meninas apresentam uma prevalência 32% menor do que os meninos, calculado como (1-0,68)x100.

De forma geral

Y: variável resposta (Ex: desnutrição)

X: variável explicativa ou de confusão (Ex: sexo)

	Variável Y		
Variável X	Y_1	Y_0	Total (%)
X_1	a	B	n_1 (100
X_0	c	D	n_0 (100
Total	m_1	m_2	n (100)

p= prevalência de Y_1= m_1/n

p_1= prevalência de $Y_1|x_1$= a/n_1

p_0= prevalência de $Y_1|x_0$= c/n_0

rp= razão de prevalências= p_1/p_0; dp=diferença de prevalências= p_1-p_0

Considere os dados abaixo

Distribuição de indivíduos segundo presença de obesidade e consumo de chocolate. Estocolmo, Suécia. Ano 2009.

Consumo de chocolate	Obesidade (> 30Kg/m^2)		
	Sim	Não	Total
Mais de 1 vez por semana	128	625	753
Até 1 vez por mês	68	353	421
Total	196	978	1174

Fonte: dados adaptados de Janszky I, Mukamal KJ, Ljung R, et al. Chocolate consumption and mortality following a first acute myocardial infarction: The Stockholm Heart Epidemiology Program. Journal of Internal Medicine 2009; 266: 248-257.

a) Calcule a prevalência de obesidade entre pessoas que consomem chocolate até 1 vez por mês.
b) Calcule a prevalência de obesidade entre pessoas que consomem chocolate mais de uma vez por semana.
c) Calcule a razão de prevalências.
d) Interprete a razão de prevalências. Você diria que a obesidade está associada ao consumo de chocolate? Justifique.

Resposta

prevalência de obesidade entre pessoas que consomem chocolate até 1 vez por mês.

$$\mathbf{\frac{68}{421} = 0,162}$$

prevalência de obesidade entre pessoas que consomem chocolate mais de uma vez por semana

$$\mathbf{\frac{128}{753} = 0,170}$$

a razão de prevalências

$$\mathbf{\frac{\frac{68}{421}}{\frac{128}{753}} = \frac{753x68}{421x128} = 0,950}$$

Conclusão algébrica: 0,95 é diferente de 1

Consumir menos chocolate protege em relação à obesidade.
Prevalência de obesidade entre pessoas que consomem menos chocolate é 0,95 vezes a prevalência de obesidade entre pessoas que consomem mais chocolate

Prevalência de obesidade entre pessoas que consomem menos chocolate é 5% menor do que a prevalência de obesidade entre pessoas que consomem mais chocolate

Razão de riscos ou risco relativo ou razão de incidências

Estudo de incidência

Distribuição de pessoas segundo hábito de fumar e morte em 5 anos por DIC. Local X. Ano Y

	Morte em 5 anos por DIC		
Fumar	Sim	Não	Total
Sim	208	850	1058
Não	264	1467	1731
Total	472	2317	2789

Fonte: dados hipotéticos.

r= 472/2789= 0,17 = 17%

r_1=208/1058= 0,20= 20%

r_0=264/1731= 0,15=15%

rr=0,20/0,15= 1,33

ra= 0,20- 0,15= 0,05= 5%

Se a razão de incidências ou RR for igual a 1 ou a diferenças de incidências for igual a 0 então se diz que as variáveis não estão associadas. Na inferência estatística é possível testar se o valor observado da razão de incidências vem de uma população com parâmetro igual a 1.

Interpretação como medida de efeito:

A incidência de óbitos entre fumantes é 1,33 vezes a incidência de óbitos entre não fumantes.

A incidência de mortes parece ser maior entre as pessoas que fumam. Os fumantes apresentam uma incidência 33% maior do que os não fumantes.

Pela diferença diz-se que 5% dos óbitos excedentes são devidos ao fumo. É possível dizer que 5% dos óbitos poderiam ser evitados na ausência do fumo.

De forma geral

Y: variável resposta
X: variável explicativa ou de confusão

	Variável Y		
Variável X	Y_1	Y_0	Total (%)
X_1	a	B	n_1 (100
X_0	c	D	n_0 (100
Total	m_1	m_2	n (100)

r= incidência de Y1= m_1/n
r_1= incidência de Y1 entre os expostos (x_1)= a/n_1
r_0= incidência de Y1 entre os não expostos (x_0)= c/n_0
r_i= razão de incidências= r_1/r_0
d_i= diferença de incidências= r_1-r_0

incidência risco

r_1 r_0 r_1/r_0 r_1-r_0

r_i=rr=**razão de riscos=risco relativo**=r_1/r_0

d_i= ra= **risco atribuível**= r_1-r_0

Investigação de toxinfecção alimentar

	Toxiinfecção		
Tomou sorvete de baunilha	Sim	Não	Total
Sim	43	11	54
Não	3	18	21
Total	46	29	75

Fonte: Epi Info, 2000.

Calcule

a) a incidência global ou taxa de ataque global
b) incidência entre quem tomou sorvete
c) incidência entre quem não tomou sorvete
d) o risco relativo ou a razão de incidências
e) Você diria que tomar sorvete de baunilha consistiu fator de risco para toxiinfecção alimentar?

Resposta

incidência global ou taxa de ataque global

$$\frac{46}{75} = 0,613$$

incidência entre quem tomou sorvete

$$\frac{43}{54} = 0,796$$

incidência entre quem não tomou sorvete

$$\frac{3}{21} = 0,143$$

a razão de incidências ou razão de riscos ou risco relativo

$$\frac{\frac{43}{54}}{\frac{3}{21}} = \frac{21x43}{3x54} = 5,6$$

Conclusão algébrica: 5,6 é diferente de 1 portanto é provável que exista associação entre as variáveis. Consumir o sorvete é fator de risco para a doença (toxiinfecção). O risco de doença entre pessoas que consumiram o sorvete é 5,6 vezes o risco da doença entre os que não consumiram.

O risco de doença entre pessoas que consumiram o sorvete é 460% maior que risco da doença entre os que não consumiram.

Razão de odds ou odds ratio

Estudo do tipo caso-controle

Odds ratio

Odds e probabilidade – duas formas diferentes de quantificar incertezas

Supor que durante um jogo de basquete um jogador acerta a cesta 2 vezes em 5 tentativas. Chamando $\hat{p}$ (p chapéu) de probabilidade de acerto tem-se que $\hat{p} = \frac{2}{5}$ = 0,4 ou 40% e a probabilidade de erro, $\hat{q} = \frac{3}{5}$ = 0,6 ou 60%.

Considerando-se que a probabilidade de acerto ou de erro = p+q= 1; então $\hat{q} = 1 - \hat{p}$.

Odds ratio

Define-se *odds* como a razão entre a probabilidade de acerto e a probabilidade de erro, ou seja, $\frac{p}{1-p}$.

No exemplo acima, o *odds* a favor de acerto é $\frac{p}{1-p} = \frac{\frac{2}{5}}{\frac{3}{5}} = \frac{2x5}{3x5} = \frac{2}{3} = 0,67$ ou 0,67:1 (0,67 acertos para 1 erro).

Estudo do tipo caso-controle

Os dados a seguir são de um estudo sobre câncer de esôfago e consumo de álcool. Local X. Ano Y.

Condição	Consumo médio de álcool (g/dia)		Total
	80 e +	0-79	
Casos	96	104	200
Controles	109	666	775
Total	205	770	975

Fonte: Tuyns et al. 1977.

(entre expostos) odds a favor de casos entre consumidores de 80 e + g/dia: $\frac{96}{205} : \frac{109}{205} = \frac{96}{109} = 0,88$

(entre não expostos) odds a favor de casos entre consumidores de 0-79g/dia: $\frac{104}{770} : \frac{666}{770} = \frac{104}{666} = 0,16$

$$\text{odds ratio: } \frac{96}{109} : \frac{104}{666} = \frac{96x666}{109x104} = 5{,}6$$

Se a razão de odds for igual a 1 então se diz que as variáveis não estão associadas. Na inferência estatística é possível testar se o valor observado vem de uma população com parâmetro igual a 1. Até agora estamos construindo a estatística que permite verificar associação.

Interpretação:
A força de morbidade de câncer de esôfago entre consumidores de 80 e + g/dias de bebida alcoólica é 5,6 vezes a força de morbidade entre os que consomem de 0 a 79g/dia.

Em casos especiais, o *odds ratio* pode ser um bom estimador do risco (quando a doença de estudo é rara).

De forma geral
Y: variável resposta
X: variável explicativa ou de confusão

	Variável Y		
Variável X	Y_1	Y_0	Total (%)
X_1	a	B	n_1 (100
X_0	c	D	n_0 (100
Total	m_1	m_2	n (100)

odds a favor de Y_1:

na categoria X_1= $(a/n_1)\div(b/n_1)$

na categoria X_0= $(c/n_0)\div(d/n_0)$

odds ratio: $[(a/n_1)\div(b/n_1)]\div[(c/n_0)\div(d/n_0)] = \frac{a/b}{c/d} = \frac{ad}{bc}$

Exercício
Os dados são de um estudo sobre consumo de pimenta e câncer gástrico, realizado no México. Ano Y.

Consumo de pimenta	Casos	Controles	Total
Sim	211	607	818
Não	9	145	154
Total	220	752	972

Fonte: X.

a) calcule o *odds* a favor de casos entre pessoas que consomem pimenta.
b) calcule o *odds* a favor de casos entre pessoas que não consomem pimenta.
c) calcule a razão dos *odds* (*odds ratio*).
d) Discuta os resultados sobre possível associação entre as variáveis.

Resposta
odds a favor de casos entre pessoas que consomem pimenta

$$\frac{\frac{211}{818}}{1-\frac{211}{818}} = \frac{\frac{211}{818}}{\frac{607}{818}} = \frac{211}{607} = 0{,}347$$

Consumo de pimenta	Casos	Controles	Total
Sim	211	607	818
Não	9	145	154
Total	220	752	972

odds a favor de casos entre pessoas que não consomem pimenta

$$\frac{\frac{9}{154}}{1-\frac{9}{154}} = \frac{\frac{9}{154}}{\frac{145}{154}} = \frac{9}{145} = 0,062$$

Razão de odds ou odds ratio (razão de razões de odds)

$$\frac{odds\ de\ doença\ entre\ expostos}{odds\ de\ doença\ entre\ não\ expostos} = \frac{\frac{211}{607}}{\frac{9}{145}} = \frac{145x211}{9x607} = \mathbf{5,6}$$

Conclusão:
Interpretação:
A força de morbidade de câncer de esôfago entre consumidores de pimenta é 5,6 vezes a força de morbidade entre os não consumidores de pimenta.

QUI-QUADRADO DE PEARSON

É uma estatística que permite verificar se existe ou não associação entre duas variáveis qualitativas.

Os exemplos são retirados de BUSSAB, Wilson de O; MORETTIN, Pedro A. *Estatística básica.* 5ª Ed. São Paulo: Saraiva, 2004.

X - curso universitário e
Y – sexo do aluno

Questão: sexo do indivíduo influi na escolha do curso?

Situação 1

Curso	Masculino (A)	Feminino	Total
	n	n	n
Economia (B)	24	36	60
Administração	16	24	40
Total	40	60	100

Definição de independência:

A – Ser do sexo masculino;
B – Estar cursando economia.

A e B são independentes se P(A e B) = P(A) x P(B).

P(A e B) = Probabilidade (ser homem e estar cursando Economia)

P(A e B) = $\frac{24}{100} = 0{,}24$

P(A) = $\frac{40}{100} = 0{,}4$ P(B) = $\frac{60}{100} = 0{,}6$

Como $\frac{24}{100} = \frac{40}{100} x \frac{60}{100}$, então A e B são independentes e, portanto não existe associação.

Notar que no caso de independência, as proporções de escolha dos cursos não diferem segundo sexo do estudante

Curso	Masculino (A)		Feminino		Total	
	n	proporção	n	proporção	n	proporção
Economia (B)	24	0,6	36	0,6	60	0,6
Administração	16	0,4	24	0,4	40	0,4
Total	40	1	60	1	100	1

Situação 2

Curso	Masculino (B) n	Feminino n	Total n
Física (A)	100 (a)	20 (b)	120
Ciências Sociais	40 (c)	40 (d)	80
Total	140	60	200

Notar que as proporções de escolha dos cursos diferem segundo sexo do estudante e a distribuição de alunos em cada curso, segundo sexo não é a mesma, sexo e curso podem estar associados

Curso	Masculino		Feminino		Total	
	n	proporção	n	proporção	n	proporção
Física	100	0,7	20	0,3	120	0,6
Ciências Sociais	40	0,3	40	0,7	80	0,4
Total	140	1	60	1	200	1

Pearson propôs uma estatística que compara os valores observados de uma distribuição conjunta observada com valores esperados calculados a partir da imposição que as variáveis sejam independentes.

Se a variável sexo não fosse associada à escolha do curso, quantos indivíduos seriam esperados em Física, entre os homens?

Aplicar a proporção marginal utilizando o raciocínio da regra de três:

x (valor desconhecido) estará para 140 assim como 120 está para 200

$\frac{x}{140} = \frac{120}{200}$ assim, $x = \frac{140x120}{200} = 84$

O mesmo raciocínio é feito para cada casela da distribuição conjunta.

$\frac{x}{140} = \frac{80}{200}$ assim, $x = \frac{140x}{200} = 56$

$\frac{x}{60} = \frac{120}{200}$ assim, $x = \frac{60x1}{200} = 36$

$\frac{x}{60} = \frac{80}{200}$ assim, $x = \frac{60x80}{200} = 24$

Curso	Masculino (B) n	Feminino n	Total n
Física (A)	100 (a)	20 (b)	120
Ciências Sociais	40 (c)	40 (d)	80
Total	140	60	200

Resumindo

Curso	Sexo	Valor Esperado sob a condição de independência
Física	Masculino (a)	$\frac{120}{200}x140 = 84$
Física	Feminino (b)	$\frac{120}{200}x60 = 36$
Ciências Sociais	Masculino (c)	$\frac{80}{200}x140 = 56$
Ciências Sociais	Feminino (d)	$\frac{80}{200}x60 = 24$

Tabela de **frequências esperadas**, sob a condição de independência.

Curso	Masculino n	Feminino n	Total n
Física	84	36	120
Ciências Sociais	56	24	80
Total	140	60	200

Estatística qui quadrado de Pearson

Valores observados O	Valores esperados E	(O-E)	$(O-E)^2$	$\frac{(O-E)^2}{E}$
100	84	16	256	3,048
40	56	-16	256	4,571
20	36	-16	256	7,11
40	24	16	256	10,667
			Qui-quadrado=	25,397

O Qui-quadrado é obtido somando-se a diferença ao quadrado entre as frequências observadas e as esperadas, dividido pelas frequências esperadas.

$$\boxed{\chi^2 = \sum \frac{(\mathrm{O}-\mathrm{E})^2}{\mathrm{E}}}$$

Se o Qui-quadrado for igual a zero, então não existe associação entre as variáveis

Se o qui quadrado for igual a 0 então se diz que as variáveis não estão associadas ou que existe independência entre elas. Na inferência estatística é possível testar se o valor observado vem de uma população com parâmetro igual a zero. Até agora estamos construindo a estatística que permite verificar associação.

Como o qui quadrado indica somente se existe ou não associação, é necessário calcular um coeficiente de associação que ajuda a verificar a força de associação.

Coeficiente de associação de Yule

Coeficiente de associação de Yule – permite investigar a força (magnitude) da associação

$$Y = \frac{a.d - b.c}{a.d + b.c} \text{ , onde: } \quad -1 \leq Y \leq +1$$

100 a b 20
40 c d 40

$$Y = \frac{100x40 - 20x40}{100x40 + 20x40} = +0{,}67$$

Exemplo:
Distribuição de recém-nascidos acometidos de síndrome de desconforto idiopático grave segundo condição de sobrevivência e peso ao nascer (g).

Peso ao nascer	Óbito		Sobrevivente		Total	
	n	%	n	%	n	%
Baixo peso (<2500)	24	64,9	13	35,1	37	100
Não baixo peso (2500 e mais)	3	23,1	10	76,9	13	100
Total	27	54,0	23	46,0	50	100

Fonte: Hand DJ et al. A handbook of small data sets. Chapman&Hall, 1994.

$\frac{x}{27} = \frac{37}{50}$

$x = \frac{27 \times 37}{50}$

Cálculo do qui-quadrado de Pearson

Valores observados O	Valores esperados E	(O-E)	$(O-E)^2$	$\frac{(O-E)^2}{E}$
24 x	19,98	4,02	16,16	0,809
3	7,02	-4,02	16,16	2,302
13	17,02	-4,02	16,16	0,949
10	5,98	4,02	16,16	2,702

Qui-quadrado=6,762

$$Y = \frac{24x10 - 3x13}{24x10 + 3x13} = \frac{240 - 39}{240 + 39} = \frac{201}{279} = +0{,}72$$

Portanto pode-se dizer que a situação de sobrevivência dos recém-nascidos pode estar associada ao peso ao nascer porque o qui-quadrado é diferente de zero. Pelo valor do coeficiente de Yule pode-se dizer que a associação é forte. Pode-se notar pela tabela a proporção de recém-nascidos que vão a óbito entre os recém-nascidos de baixo peso é maior do que a proporção de óbitos entre os que não nasceram com baixo peso. Isto indica que o peso ao nascer é importande para explicar as proporções de ocorrência de óbito indicando existência de associação entre os eventos.