Medidas de Associação de dados Qualitativos Nominais

Profa Ana Amélia Benedito Silva

aamelia@usp.br

Medidas de Associação de variáveis qualitativas nominais

- 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: n indivíduos são observados e classificados segundo 2 variáveis X e Y

X	Y		Total
	Y_1 (sim)	Y_0 (não)	
X_1 (sim)	A	B	n_1
X_0 (não)	C	D	n_0
Total	m_1	m_2	n

Razão de Prevalências

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

	Y		
X	Y_1 (sim)	Y_0 (não)	Total
X_1 (sim)	A	B	n_1
X_0 (não)	C	D	n_0
Total	m_1	m_2	n

Razão de Prevalências

São apresentados dados sobre o estado nutricional de 1226 crianças brasileiras de 2 anos separadas por sexo.

	sexo		
estado nutricional	masculino	feminino	Total
desnutridas	29	20	49
normais	574	603	1177
Total	603	629	1226

Razão de Prevalências

estado nutricional	sexo		Total
	masculino	feminino	
desnutridas	29	20	49
normais	574	603	1177
Total	603	623	1226

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

Razão de Prevalências

estado nutricional	sexo		Total
	masculino	feminino	
desnutridas	29	20	49
normais	574	603	1177
Total	603	623	1226

Prevalência de desnutrição segundo sexo:

PM = prevalência de desnutridas em M = 29/603= 0,048

PF = prevalência de desnutridas em F = 20/623= 0,032

Razão de Prevalências

estado nutricional	sexo		Total
	masculino	feminino	
desnutridas	29	20	49
normais	574	603	1177
Total	603	623	1226

Razão de Prevalências de desnutrição segundo sexo:

RP = 0,048/0,032= 1,5

Diferença de Prevalências de desnutrição segundo sexo:

DP = 0,048-0,032=0,016

Razão de Prevalências

P = prevalência de crianças desnutridas = 4%
PM = prevalência de meninos desnutridos = 4,8%
PF = prevalência de meninas desnutridas = 3,2%
RP = razão de prevalências = 1,56 (ou 156%)
DP = diferença de prevalências = 1,8%

- A prevalência de desnutrição é maior entre os meninos.
- A prevalência dos meninos é 1,5 vezes (uma vez e meia) a prevalência das meninas.
- A prevalência dos meninos é 50% maior que a prevalência entre meninas, calculado como (1,5-1)x100.

Razão de Prevalências

P = prevalência de crianças desnutridas = 4%
PM = prevalência de meninos desnutridos = 4,8%
PF = prevalência de meninas desnutridas = 3,2%
RP = razão de prevalências = 1,56
DP = diferença de prevalências = 0,018

- Quanto à DP, pode-se dizer que a prevalência entre meninos excede a de meninas em 1,8% ou que a diferença entre as prevalências é de 1,8%.
- Se a RP for igual a 1 ou a DP for igual a 0, diz-se que as variáveis não estão associadas.
- Na inferência estatística é possível testar se o valor observado vem de uma população com parâmetro RP=1 ou DP = 0.

Razão de Incidências

- Estudo de incidência: n indivíduos são observados e classificados segundo 2 variáveis X e Y

	Y		
X	Y_1 (sim)	Y_0 (não)	Total
X_1 (sim)	A	B	n_1
X_0 (não)	C	D	n_0
Total	m_1	m_2	n

Razão de Incidências

R_1 = incidência de Y_1 entre os X_1 = A/n_1

R_0 = incidência de Y_1 entre os X_0 = C/n_0

RI = risco de incidência = R_1/R_0

RR = Risco relativo = R_1/R_0 = RI

Ra = diferença entre riscos = $R_1 - R_0$

	Y		
X	Y_1 (sim)	Y_0 (não)	Total
X_1 (sim)	A	B	n_1
X_0 (não)	C	D	n_0
Total	m_1	m_2	n

Razão de Incidências

Considere um estudo em que se deseje verificar o risco de pacientes portadores de hipertensão vir a ter um infarto.

Por meio de um experimento duplo-cego, 263 sujeitos, sendo 114 hipertensos, foram acompanhados por 4 anos para detectar a eventual ocorrência de infarto.

Razão de Incidências

Fator de risco (hipertensão)	infarto		Total
	sim	não	
presente	60	54	114
ausente	28	121	149
Total	88	175	263

Razão de Incidências

Fator de risco (hipertensão)	infarto		Total
	sim	não	
presente	60	54	114
ausente	28	121	149
Total	88	175	263

R_1 = incidência de infarto entre sujeitos hipertensos

$$R_1 = 60/114 = 0{,}5263\ (52{,}63\%)$$

R_0 = incidência de infarto entre sujeitos não-hipertensos

$$R_0 = 28/149 = 0{,}1879\ (18{,}79\%)$$

Razão de Incidências

Fator de risco (hipertensão)	infarto		Total
	sim	não	
presente	60	54	114
ausente	28	121	149
Total	88	175	263

RR = Risco Relativo = [60/(60+54)] / [28/(28+121)] = 2,8

Ra = Diferença entre riscos = R1–R0 = 0,5263-0,1879 = 0,3384

Razão de Incidências

RR = 2,8

Significado: A incidência relativa de infarto dentre os hipertensos é quase 3 vezes maior do que dentre os não-hipertensos.

Ra = 33,84%

Significado: A diferença entre incidências significa que 33,84% dos sujeitos infartados são hipertensos.

Pode-se dizer que 33,84% dos infartos poderiam ser evitados se os sujeitos hipertensos se tratassem.

Razão de Incidências

- Se a razão de incidências ou RR for igual a 1 ou a diferenças de incidências Ra for igual a 0, então diz-se 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.

- O RR indica a chance de alguém, que tenha sido exposto a um fator de risco para uma certa doença, venha a desenvolver esta doença, quando comparado com alguém que não foi exposto ao fator de risco.

Exercício

- Foi relatado um surto de doença diarreica ocorrido entre participantes de um Encontro de 75 estudantes num colégio.
- 46 estudantes apresentaram diarreia, sendo que 43 deles tomaram sorvete de baunilha. Dentre os 29 que não tiveram diarreia, apenas 11 tomaram sorvete de baunilha.
- Deseja-se verificar possíveis associações entre o sorvete e a diarreia.

Exercício

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

Calcular:

a) A incidência de diarréia entre estudantes que tomaram sorvete de baunilha.

b) A incidência de diarréia entre estudantes que não tomaram sorvete de baunilha.

c) A razão de incidências ou Risco relativo de óbito

d) O RR sugere que as variáveis são associadas? Justifique

Exercício

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

Peso ao nascer	Óbito	Sobrevida	Total
Baixo peso (<2500)	24	13	37
Não baixo peso (2500 e mais)	3	10	13
Total	27	23	50

Calcular:

a) A incidência de óbitos entre crianças com baixo peso

b) A incidência de óbitos entre crianças sem baixo peso

c) A razão de incidências ou Risco relativo de óbito

d) O RR sugere que as variáveis são associadas? Justifique

Exercício

Peso ao nascer	Óbito	Sobrevida	Total
Baixo peso (<2500)	24	13	37
Não baixo peso (2500 e mais)	3	10	13
Total	27	23	50

R1 = incidência de óbitos com baixo peso = 24/37 = 0,65

R0 = incidência de óbitos sem baixo peso = 3/13 = 0,23

RR = [(24/37) / (3/13)] = 0,65/0,23 = 2,81

Razão de odds ou Odds Ratio

- n indivíduos são observados e classificados segundo 2 variáveis X e Y

	Y		
X	Y_1 (sim)	Y_0 (não)	Total
X_1 (sim)	A	B	n_1
X_0 (não)	C	D	n_0
Total	m_1	m_2	n

Odds e probabilidade

- Suponha que durante um jogo de basquete um jogador acerta 2 vezes em 5 tentativas.
- Seja p a probabilidade de acerto e q a probabilidade de erro:

  $p = 2/5 = 0,4$ ou 40%

  $q = 3/5 = 0,6$ ou 60%
- Como $p+q= 1$ então $q=1-p$

Odds e probabilidade

Odds – chance

elemento importante na regressão logística (técnica de análise multivariada)

- Define-se odds como sendo a razão entre p e q, ou seja, razão entre 2 probabilidades

$$Odds = p/q = p/(1-p)$$

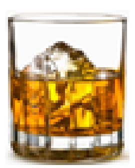
- No exemplo do basquete odds a favor de acerto é :

$Odds = p/(1-p)=(2/5)/(3/5) = (2x5)/(3x5) =2/3 = 0,67$

Interpretação: 0,67:1 (0,67 acertos para 1 erro)

Estudo do tipo caso-controle

- Suponha um estudo sobre câncer de esôfago e consumo de álcool.

~ 12 g de álcool puro				
330 ml de cerveja 	**=**	**100 ml de vinho** 	**=**	**30 ml de destilado** 

Estudo do tipo caso-controle

Câncer de esôfago e consumo de álcool.

Consumo médio de álcool por dia	Casos (tem câncer)	Controles (não tem câncer)	Total
80g ou +	96	109	205
até 79g	104	666	770
Total	200	775	975

Odds a favor de casos entre expostos a fator de risco (>=80g) = (96/205)/(109/205) = 96/109 = 0,88

Odds a favor de casos entre não-expostos a fator de risco (<=79g) = (104/770)/(666/770) = 104/666 = 0,16

OR = (0,88/0,16) = 5,6

Estudo do tipo caso-controle

Conclusão:

A chance de ter câncer de esôfago dentre consumidores de >=80 g/dia de bebida alcoólica é 5,6 maior que dentre os que consomem de <=79 g/dia.

Se OR = 1 , não há diferença de risco entre os grupos

Se OR ≠ 1 , há diferença de risco entre os grupos

Odds Ratio (OR)

- Suponha que um pesquisador disponha de 1000 pacientes, 160 com doenças respiratórias e 840 sem doenças respiratórias.
- O fator de risco é o hábito de fumar dos pais.
- Calcule o OR

Odds Ratio (OR)

Fator de risco	Doença respiratória presente (casos)	Doença respiratória ausente (controles)	TOTAL
Presente (pais fumam)	120	30	150
Ausente (pais não fumam)	40	810	850
TOTAL	160	840	1000

Odds a favor dos casos entre expostos ao fumo $= \dfrac{\frac{120}{150}}{\frac{30}{150}} = 4{,}0$

Odds a favor dos casos entre não expostos ao fumo $= \dfrac{\frac{40}{850}}{\frac{810}{850}} = 0{,}049$

Odds Ratio (OR)

Fator de risco	Doença respiratória presente (casos)	Doença respiratória ausente (controles)	TOTAL
Presente (pais fumam)	120	30	150
Ausente (pais não fumam)	40	810	850
TOTAL	160	840	1000

ODDS RATIO = $\frac{4}{0,049}$ **= 81,6**

Significado: A chance de ter Doença Respiratória tendo pais que fumam é 81,6 vezes maior do que ter Doença Respiratória tendo pais que não fumam.

Qui-quadrado

Mede a dependência entre 2 variáveis qualitativas X e Y

Exemplo: X - curso universitário e Y – sexo do aluno

Questão: o sexo do indivíduo influi na escolha do curso?

	SEXO		
CURSO	Masculino	Feminino	Total
Economia	A	B	n_1
Administração	C	D	n_0
Total	m_1	m_2	n

Qui-quadrado

- Mede a dependência entre 2 variáveis qualitativas X e Y
- Exemplo: X - curso universitário e Y – sexo do aluno
- Questão: o sexo do indivíduo influi na escolha do curso?

	SEXO		
CURSO	Masculino	Feminino	Total
Economia	A	B	n_1
Administração	C	D	n_0
Total	m_1	m_2	n

Qui-quadrado

SITUAÇÃO 1 -

Curso	Masculino	Feminino	Total
	n	n	n
Economia	24	36	60
Administração	16	24	40
Total	40	60	100

Curso	Masculino		Feminino		Total	
	n	proporção	n	proporção	n	proporção
Economia	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

As proporções de escolha dos cursos não difere segundo sexo do estudante

Qui-quadrado

SITUAÇÃO 2

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

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

A distribuição de alunos em cada curso, segundo sexo não é a mesma. Sexo e curso podem estar associados.

Qui-quadrado

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

  Esperaríamos : (120/200)*140 = 84

- Se a variável sexo **não fosse associada** à escolha do curso, quantos indivíduos **esperaríamos** em Ciências Sociais, entre os homens?

  Esperaríamos: : (80/200)*140 =56

- Se a variável sexo **não fosse associada** à escolha do curso, quantos indivíduos **esperaríamos** em Física, entre as mulheres?

  Esperaríamos : (120/200)*60 = 36

- Se a variável sexo **não fosse associada** à escolha do curso, quantos indivíduos **esperaríamos** em Ciências Sociais, entre as mulheres?

  Esperaríamos: (80/200)*60 =24

Qui-quadrado

Tabela esperada, 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

Qui-quadrado

Valores observados O	Valores esperados E	(O-E)	$(O\text{-}E)^2$	$\frac{(O-E)^2}{E}$
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 freqüências observadas e as esperadas, dividido pelas freqüências esperadas

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

Se o Qui-quadrado for igual a zero, então não existe associação entre as variáveis. Quanto maior o valor do Qui-quadrado, maior a chance de existir associação entre as variáveis, entretanto, o Qui-quadrado não mede força de associação.

Obrigada