

Revista de Saúde Pública

JOURNAL OF PUBLIC HEALTH

Razões de proporções e uso do método delta para intervalos de confiança em regressão logística

Ratio of proportions and the use of the delta method for confidence interval estimation in logistic regression

Nelson F. de Oliveira, Vilma S. Santana e Antonio Alberto Lopes

Instituto de Matemática da Universidade Federal da Bahia. Salvador, BA - Brasil (N.F.O.);

Instituto de Saúde Coletiva da Universidade Federal da Bahia. Salvador, BA - Brasil (V.S.S.);

Departamento de Medicina da Universidade Federal da Bahia. Salvador, BA - Brasil (A.A.L.)

OLIVEIRA, Nelson F. de, Razões de proporções e uso do método delta para intervalos de confiança em regressão logística. *Rev. Saúde Pública*, 31 (1): 90-9, 1997.

Razões de proporções e uso do método delta para intervalos de confiança em regressão logística

Ratio of proportions and the use of the delta method for confidence interval estimation in logistic regression

Nelson F. de Oliveira, Vilma S. Santana e Antonio Alberto Lopes

Instituto de Matemática da Universidade Federal da Bahia. Salvador, BA - Brasil (N.F.O.); Instituto de Saúde Coletiva da Universidade Federal da Bahia. Salvador, BA - Brasil (V.S.S.); Departamento de Medicina da Universidade Federal da Bahia. Salvador, BA - Brasil (A.A.L.)

Resumo

Objetivo Apresentar uma aplicação da regressão logística para estimar razões de prevalência ou risco relativo, e o uso do método delta para a estimativa de intervalos de confiança.

Métodos Utilizou-se o método delta, visto que esta técnica é adequada para estimar variâncias de funções não lineares de variáveis aleatórias, do tipo razão de proporções. O método baseia-se na expansão da função em série de Taylor, provendo uma aproximação para o cálculo das variâncias. Para estimar as razões de proporções e as respectivas variâncias, que permitem a estimação dos intervalos de confiança, desenvolveu-se programa utilizando-se o módulo matricial do SAS. A aplicação prática do modelo proposto é feita utilizando-se dados de um estudo transversal com uma amostra de 611 mulheres, para testar a hipótese de que a falta de compartilhamento do trabalho doméstico associa-se com altos escores de sintomas psicológicos medidos através de um questionário validado. As razões de prevalência calculadas através da regressão logística, e os intervalos de confiança estimados pelo método delta foram comparadas com os resultados obtidos pela técnica de Mantel-Haenszel.

Resultados As estimativas pontuais das razões de prevalência brutas, cruas ou ajustadas, obtidas através da regressão logística, foram semelhantes às encontradas na análise tabular. Também os limites dos intervalos de confiança para as razões de prevalência através do método delta foram praticamente iguais em relação aos encontrados através da técnica de Mantel-Haenszel.

Conclusões Os resultados apóiam o uso do método delta para estimar intervalos de confiança para razões de proporção. Esse método deve ser visto como uma alternativa para situações em que o efeito em estudo (variável dependente) não é raro, situação onde a "odds ratio" se afasta das medidas epidemiológicas caracterizadas como razão de proporções.

Regressão logística. Intervalo de confiança. Risco relativo. Método delta.

Abstract

- Objective** *To present an application of logistic regression modelling to estimate ratios of proportions, such as prevalence ratio or relative risk, and the Delta Method to estimate confidence intervals.*
- Method** *The Delta Method was used because it is appropriate for the estimation of variance of non-linear functions of random variables. The method is based on Taylor's series expansion and provides a good approximation of variance estimates. A computer program, utilizing the matrix module of SAS, was developed to compute the variance estimates. A practical demonstration is presented with data from a cross-sectional study carried out on a sample of 611 women, to test the hypothesis that the lack of housework sharing is associated with high scores of psychological symptoms as measured by a validated questionnaire.*
- Results** *Crude and adjusted prevalence ratio estimated by logistic regression were similar to those estimated by tabular analysis. Also, ranges of the confidence intervals of the prevalence ratio according to the Delta Method were nearly equal to those obtained by the Mantel-Haenszel approach.*
- Conclusions** *The results give support to the use of the Delta Method for the estimation of confidence intervals for ratios of proportions. The method should be seen as an alternative for situations in which the need to control a large number of potential confounders limits the use of stratified analysis.*
- Logistic regression. Confidence intervals. Relative risk. Delta method.**

INTRODUÇÃO

O uso de modelagem matemática, especialmente a regressão logística, tem se tornado uma prática quase rotineira na análise epidemiológica. Aparentemente, isso se relaciona com a popularidade de pacotes estatísticos de utilização muitas vezes fácil e rápida, que permitem o manejo de múltiplas variáveis simultaneamente. Todavia, enquanto nos estudos de caso-controle a estimativa da "odds ratio" (OR) é obtida pela simples exponenciação do coeficiente de regressão da variável independente de interesse, nos estudos transversais e coortes com populações fixas, as medidas de associação de maior interesse epidemiológico são, respectivamente, a razão de prevalências (RP) e a razão de incidência cumulativa¹⁴ (risco relativo ou RR). Este é provavelmente o motivo principal para a crescente apresentação de "odds ratios" de prevalência (ORP) ou "odds ratio" de incidência (ORI) como medidas de associação em estudos que não os de caso-controle.

As ORs têm tido o seu significado epidemiológico per se questionado, apontando-se para o seu valor interpretativo apenas quando se aproxima das estimativas de razões de proporções ou de funções de densidade de incidências⁸. Vários estudos têm apontado as limitações das razões de produtos cruzados, as

ORs, como estimadores de RPs. Por exemplo, aproximação matemática entre RPs e ORs ocorre apenas em certas situações específicas, como a raridade do evento predito (doença) entre expostos e não expostos, e estabilidade da prevalência da exposição^{11, 12}. Exceção para os estudos de caso-controle, coorte por amostragem baseada na densidade de incidência, nos quais os controles são selecionados do conjunto de não-casos a cada vez que ocorre um caso⁶. Isto tem levado a que alguns autores^{4, 11} recomendem o uso de OR apenas nas situações onde esta medida pode ser tomada como uma aproximação não viesada de razões de densidade de incidência, embora outros^{3, 8} indiquem o uso de OR para estudos transversais. Discussão mais ampla desta questão vem sendo recentemente desenvolvida por Lee e Chia⁹, Axelsson e col.¹, Stromberg¹⁵ e Lee¹⁰.

Enquanto é possível, com alguns cálculos, se estimar as razões de proporções com base nos coeficientes de regressão logística, mesmo quando o ajustamento para variáveis de confusão se faz necessário, a estimação de intervalos de confiança para estas medidas requer procedimento de difícil operacionalização. Neste trabalho, usa-se o modelo de regressão logística para estimar razões de proporções e desenvolve-se, baseado no método delta², um procedimento para estimação de variâncias e, con-

seqüentemente, intervalos de confiança para as razões de proporções. Este procedimento envolve cálculo matricial e pode ser implementado mediante programas computacionais que disponham de recursos para manipulação de matrizes. Deve ser notado que este método não se aplica para os estudos de função de risco (hazard function) ou de densidade de incidência, quando a medida de associação não é uma razão de proporções ou de probabilidades.

O objetivo do presente artigo é apresentar uma aplicação da regressão logística para estimar razões de prevalências ou risco relativo, e o uso do método delta² para a estimativa de intervalos de confiança.

APLICAÇÃO DA MODELAGEM LOGÍSTICA EM ANÁLISE EPIDEMIOLÓGICA

A regressão logística tem a sua aplicação mais imediata para a estimativa de ORs. Isso se deve ao fato de que com o antilogaritmo do coeficiente de regressão (e^β) de uma variável independente obtém-se uma estimativa não viesada da OR para a associação entre essa variável e a dependente. Assim, para a associação entre uma variável dependente D e uma variável independente E, considerando-se todas as variáveis como dicotômicas e definidas como de desenho, "dummy variables", com a seguinte codificação: presença do fator=1 e ausência do fator=0, o modelo para o logito da probabilidade de ter o efeito D ($D = 1$) dada a exposição E ($E = 1$), na ausência de interação, com erro aleatório e , pode ser escrito como

$$\text{logito} [\Pr(D = 1)] = \beta_0 + \beta_1 E + e \quad (1)$$

$$\text{onde } \text{logito} [\Pr(D = 1)] = \ln \frac{\Pr(D = 1)}{1 - \Pr(D = 1)} .$$

Assim, a estimativa da OR *não ajustada ou bruta*, ($OR_{(D|E)}$), poderá ser calculada da seguinte forma:

$$\widehat{OR}_{(D|E)} = e^{\hat{\beta}_1} \quad (2)$$

Quando covariáveis de efeito de confusão são identificadas, na ausência de interação, é recomendado o ajustamento da medida da associação principal para essas covariáveis. Para estimar medidas ajustadas da OR para o efeito de confusão de covariáveis, por exemplo, da covariável C, o ajustamento advirá da manutenção dessa covariável no modelo final. Isto é

$$\text{logito} [\Pr(D = 1)] = \beta_0 + \beta_1 E + \beta_2 C + e \quad (3)$$

e a estimativa da "odds ratio" ajustada será dada pela mesma expressão em (2), com $\hat{\beta}_1$ sendo estimado através do modelo (3).

Estes são, portanto, os caminhos básicos para estimativa da medida de associação para estudos de caso-controle com resultados provenientes da regressão logística. Estas equações aplicam-se para estudos não pareados, ou seja, se prestam à situação de modelo de regressão logística não condicional, considerando-se o atendimento a todos os pressupostos desse modelo, observando-se que inferência sobre β_0 não pode ser feita a menos que se conheça a fração amostral de casos⁷.

ESTIMATIVA DE RAZÕES DE PROPORÇÕES COM REGRESSÃO LOGÍSTICA

É bem conhecido que através de modelos de regressão logística (1) é possível determinar probabilidade de ocorrência da variável dependente D, de acordo com cada categoria da variável independente⁷. Em epidemiologia, a variável dependente é comumente representada por um evento (por exemplo, doença) e a variável independente por um fator de exposição. Em um estudo de corte transversal, as proporções da variável dependente entre expostos e não expostos permitem determinar a razão de prevalência (RP). Semelhantemente, o risco relativo (RR) poderá ser determinado num estudo de coorte fixa, situação onde a proporção de indivíduos que mudam o estado de não-doença para o de doença representa o risco ou incidência cumulativa.

Na Tabela 1, do tipo 2 x 2, apresentam-se os dados para uma situação onde as variáveis relativas à associação principal são dicotômicas. As colunas representam os dois níveis da variável dependente ($D(1 = doente, 0 = não doente)$) e as linhas, os dois níveis da variável independente principal, ou de exposição ($E(1 = exposto, 0 = não exposto)$). Tomando-se os coeficientes $\hat{\beta}_0$ e $\hat{\beta}_1$ fornecidos pelo modelo de regressão logística, determina-se a probabilidade (a prevalência ou o risco cumulativo, de acordo com a situação particular) de doença entre os expostos como

$$\Pr(D = 1|E = 1) = \frac{e^{\hat{\beta}_0 + \hat{\beta}_1}}{1 + e^{\hat{\beta}_0 + \hat{\beta}_1}} \quad (4)$$

Entre os não expostos, a probabilidade de doença é dada por

$$\Pr(D = 1|E = 0) = \frac{e^{\hat{\beta}_0}}{1 + e^{\hat{\beta}_0}} \quad (5)$$

A razão de probabilidades (a razão de prevalência ou o risco relativo) é o resultado da divisão da

Tabela 1 - Equações para estimativa das probabilidades correspondentes a uma tabela 2 x 2 através de modelos logísticos (Eq. 1).
Table 1 - Equations for the estimates of the probabilities corresponding to a 2 x 2 table through logistic models (Eq. 1).

	Doente (D = 1)	Não Doente (D = 0)
Exposto (E = 1)	$Pr(D = 1 E = 1) = \frac{e^{0+1}}{1 + e^{0+1}}$	$Pr(D = 0 E = 1) = \frac{1}{1 + e^{0+1}}$
Não exposto (E = 0)	$Pr(D = 1 E = 0) = \frac{e^0}{1 + e^0}$	$Pr(D = 0 E = 0) = \frac{1}{1 + e^0}$

probabilidade do evento entre expostos pela probabilidade do evento entre não expostos, ou seja

$$Pr(D = 1|E = 1)/Pr(D = 1|E = 0) = \frac{e^{0+1}}{1 + e^{0+1}} \div \frac{e^0}{1 + e^0} \quad (6)$$

ESTIMAÇÃO DA VARIÂNCIA DA RAZÃO DE PREVALÊNCIA E INTERVALO DE CONFIANÇA

Nesta seção, para simplicidade da notação, as variáveis independentes, sejam de confusão para a associação principal ou a de exposição, serão denotadas por X_1, X_2, \dots , etc. A variável X_1 (1 = exposto, 0 = não exposto) é dicotômica, enquanto que as variáveis de confusão poderão ser de qualquer natureza. Será considerado um modelo geral com r variáveis independentes. Neste contexto, o estimador de $Pr(D = 1|X_1, X_2, \dots, X_r)$, é dado por

$$\hat{p} = Pr(D = 1|\hat{X}_1, \hat{X}_2, \dots, \hat{X}_r) = 1 + \exp^{-\sum_{i=0}^r \hat{X}_i} = \left(\hat{X}_0, \hat{X}_1, \dots, \hat{X}_r \right) \quad (7)$$

com $X_0 = 1$.

O estimador da razão de prevalência denotado por \hat{r} , é dado por

$$\hat{r} = \hat{p}(X_1 = 1)/\hat{p}(X_1 = 0)$$

ou simplificada $\hat{r} = \hat{p}_1/\hat{p}_0$. Assume-se que $\ln(\hat{r})$ tem distribuição aproximadamente normal, com média $E[\ln(\hat{r})]$ e variância $Var[\ln(\hat{r})]$. Para estimar $Var[\ln(\hat{r})]$ será empregado o método delta que se baseia no desenvolvimento, em série de Taylor, da função $\ln(\hat{r})$. De modo geral, esse método provê uma boa aproximação para médias, variâncias e covariâncias de funções não lineares de uma ou mais variáveis aleatórias. Essencialmente, o método consiste na expansão em série de Taylor da função de interesse, sendo usual tomar-se a aproximação de primeira ordem para a média e de segunda ordem para a variância. No caso de mais de uma variável, o mé-

todo envolve o cálculo de derivadas parciais da função. Usando-se a notação $g(X_1, X_2, \dots, X_r)$ para função genérica de interesse, $E(X_i) = \mu_i$ para as médias, $Var(X_i)$ para as variâncias das r variáveis e $Cov(X_i, X_j)$ para a covariância entre duas variáveis, obtém-se as seguintes aproximações.

$$E[g(X_1, X_2, \dots, X_r)] \approx g(\mu_1, \mu_2, \dots, \mu_r)$$

$$Var[g(X_1, X_2, \dots, X_r)] \approx \sum_{i=1}^r \frac{g^2}{\mu_i} \times Var(X_i) + 2 \sum_{i < j} \frac{g^2}{\mu_i \mu_j} Cov(X_i, X_j)$$

Do mesmo modo obtém-se a seguinte aproximação para a covariância entre duas funções não lineares g_1 e g_2 de variáveis aleatórias

$$Cov[g_1(X_1, X_2, \dots, X_r), g_2(X_1, X_2, \dots, X_r)] \approx \sum_{i=1}^r \frac{g_1}{\mu_i} \frac{g_2}{\mu_i} Var(X_i) + 2 \sum_{i < j} \frac{g_1}{\mu_i} \frac{g_2}{\mu_j} Cov(X_i, X_j),$$

onde o símbolo $\frac{g}{\mu_i}$ representa a derivada parcial da função g com relação a variável X_i , em cujo resultado substituiu-se X_i por μ_i . Estes resultados são aplicados no presente trabalho (veja Apêndice 1), tanto no caso multivariado como no caso univariado. Após a estimação de $Var[\ln(\hat{r})]$, pode-se determinar um intervalo de $(1 - \alpha) \times 100\%$ de confiança para a razão de prevalência, através de

$$e^{\ln(\hat{r}) \pm z_{\alpha/2} \sqrt{Var[\ln(\hat{r})]}} \quad (8)$$

onde $z_{\alpha/2}$ é o ponto da distribuição normal padronizada acima do qual há uma probabilidade de $\alpha/2$.

Para a estimação da razão de incidências cumulativas, ou risco relativo, e as suas respectivas variâncias, que permitirão a estimação dos intervalos de confiança, desenvolveu-se um programa utilizando o módulo matricial "Interactive Matrix Lan-

guage” do “SAS”¹³ (IML/SAS), apresentado no Apêndice 2. Para sua aplicação, em uma situação concreta de análise, indica-se a realização das seguintes etapas:

- 1) definir o modelo final através do processo de modelagem logística comum;
- 2) estimar os parâmetros do modelo da regressão logística (coeficientes de regressão) e a matriz de variância-covariância do vetor de parâmetros;
- 3) estimar a razão de prevalências para a associação principal baseada no modelo final definido;
- 4) especificar os vetores de dados;
- 5) estimar $\text{Var}[\ln(\hat{\rho})]$, de acordo com a expressão matricial (9) desenvolvida no Apêndice 1;
- 6) estimar o intervalo de confiança com as especificações desejadas, através de (8).

EXEMPLO EM UMA SITUAÇÃO DE ANÁLISE EPIDEMIOLÓGICA

Com dados de um estudo transversal, realizado em uma amostra aleatória de 470 domicílios de um bairro de Salvador, analisou-se a hipótese de que a falta de compartilhamento do trabalho doméstico era um fator de risco para número de sintomas psicológicos em mulheres. A população de estudo era composta de 611 mulheres com idades entre 14 e 70 anos, para as quais dispunha-se de informações sociodemográficas, número de mulheres que desenvolviam atividades domésticas na mesma família e respostas a um questionário so-

bre sintomas psicológicos, o Questionário de Morbidade Psiquiátrica de Adultos (QMPA). Para essa análise, foi definida como variável de exposição o Compartilhamento do Trabalho Doméstico (COMP), do tipo dicotômica (1 = não compartilhamento, 0 = compartilhamento) e como variável dependente, o escore do QMPA (SCOR) codificado como: 1 = escore alto (> 7 respostas positivas), 0 = escore baixo (< 7 respostas positivas). Durante a análise, identificou-se que o estado civil (ECIV) e nível de educação formal (EDC), ambas categorizadas como dicotômicas (1 = casadas ou de menor escolarização e 0 = solteiras ou de maior escolarização) eram de confusão da associação principal. Assim, o modelo final pôde ser escrito da seguinte forma:

$$\text{logito Pr(SCOR} = 1) = \beta_0 + \beta_1(\text{COMP}) + \beta_2(\text{ECIV}) + \beta_3(\text{EDC})$$

Na análise tabular, utiliza-se o Método de Mantel-Haenszel para estimação da razão de prevalências sendo que os intervalos de confiança são baseados no método de Greenland e Robins (1985)⁵, enquanto que com a análise logística emprega-se estimativas baseadas na máxima verossimilhança. Encontrou-se uma prevalência global de escores altos do QMPA de 24,60%, considerando-se, para simplicidade de análise, a amostragem como aleatória simples. Para esta magnitude de ocorrência do efeito, a OR não é um bom estimador da razão de prevalências, medida de escolha para a análise desse tipo de estudo, ensejando uma situação para o uso de parâmetros da regressão logística para a estimativa adequada.

Tabela 2 - Razões de Prevalência, Odds Ratios e Intervalos de Confiança a 95%, para a associação entre Falta de Compartilhamento do Trabalho Doméstico (COMP) e Escores altos do QMPA (SCOR) obtidas com a análise estratificada.

Table 2 - Prevalence Ratios, Odds Ratios and 95% Confidence Intervals for the association of the lack of the sharing of housework (COMP) with high scores of the QMPA (SCOR) obtained by stratified analysis.

Especificação	Nº	RP	95% IC	ORP	95% IC
Bruta	611	1,90	(1,34 2,69)	2,29	(1,49 3,48)
Estado civil					
Casadas	362	1,55	(1,00 2,41)	1,84	(1,03 3,27)
Solteiras	249	1,72	(0,95 3,12)	1,90	(0,94 3,83)
Ajustada		1,61	(1,13 2,29)	1,86	(1,12 2,90)
Educação					
Menor/elementar	329	1,47	(0,98 2,21)	1,72	(0,99 2,98)
Maior/elementar	282	2,42	(1,29 4,54)	2,83	(1,54 5,82)
Ajustada		1,75	(1,24 2,46)	2,08	(1,36 3,20)
Ajustada para estado civil e educação	611	1,56	(1,08 2,24)	1,80	(1,14 2,86)

RP - Razão de Prevalências; ORP - Odds Ratio da Prevalência; 95%IC - Intervalo de Confiança para 95%; Intervalos de confiança das RPs pelo método de Greenland e Robins (1985) e para as OR pelo método de Mantel-Haenszel. QMPA (SCOR) Questionário de Morbidade Psiquiátrica de Adultos.

De acordo com os dados da Tabela 2, evidencia-se uma associação positiva entre a falta de compartilhamento do trabalho doméstico e escores altos do QMPA, estatisticamente significativa ao nível de 5%. Como se esperava, as estimativas pontuais para as associações evidenciam um padrão regular de superestimação das ORs em comparação às RPs. Além disso, a amplitude dos intervalos de confiança (IC) correspondentes, tanto para as medidas brutas, estrato-específicas, ou ajustadas para cada uma das variáveis de confusão, separadamente ou simultaneamente, são maiores para as ORs quando comparadas às RPs. Assim, pode-se dizer que para este efeito, não raro (prevalência de 24,6%), o uso de OR em vez de RP superestima a magnitude da medida de associação. Quanto à precisão, os resultados apontam para uma maior imprecisão das ORs quando comparadas às RPs. Portanto, conclusões poderão ser distintas, de acordo com a medida de associação utilizada.

Na Tabela 3 podem ser vistos os resultados baseados nos coeficientes da regressão logística. Consistentemente, as medidas para a associação e intervalos de confiança para OR e RP diferem, seguindo o mesmo padrão observado na análise tabular. Quando se comparam os resultados obtidos para as estimativas pontuais das ORs ou RPs – estas últimas calculadas com base nas equações derivadas do modelo logístico (Tabela 1, Eq. 1) – nota-se que os valores são essencialmente iguais àqueles calculados na análise tabular. Exceção para pequenas diferenças entre as medidas ajustadas para cada uma das variáveis de confusão, separadamente ou ajuste múltiplo.

Quanto às estimativas dos intervalos de confiança para as RPs, também mostradas na Tabela 3, evidencia-se que as amplitudes obtidas com o método delta² são muito semelhantes, como era de se esperar, àquelas obtidas na análise estratificada tabular, onde o método delta também é usado para a estimativa das variâncias das ORs. Notam-se pequenas diferenças nas estimativas das RPs ajustadas para uma ou duas variáveis de confusão, quando comparadas àquelas que foram obtidas com as ORs. Portanto, pode-se sugerir o uso desse método, desde que os valores estimados, pontuais ou intervalares são essencialmente iguais aos que advêm da análise tabular. Com isto, uma mesma medida de associação poderá ser utilizada para comparação de resultados provenientes de modelagem logística e análise tabular, evitando-se desse modo o uso inadequado de OR em estudos transversais, que ocasionam superestimação das estimativas pontuais e amplificação da imprecisão evidente nos intervalos de confiança.

Nas Tabelas 4 e 5 apresentam-se os resultados obtidos com o múltiplo controle das variáveis de confusão para as ORs e RPs com a análise tabular e modelagem logística. Observou-se similaridade das estimativas obtidas através dos dois procedimentos, ambos baseados no método delta para o cálculo das variâncias. Os valores ajustados mostram pequenas diferenças uma vez que as estimativas pontuais são calculadas por diferentes métodos, Mantel- Haenszel e regressão logística.

Tabela 3 - Razões de Prevalência, Odds Ratio e Intervalos de Confiança para 95% para a associação entre Falta de Compartilhamento do Trabalho Doméstico (COMP) e Escores altos do QMPA (SCOR) baseadas na Regressão Logística.

Table 3 - Prevalence Ratios, Odds Ratio and 95% Confidence Intervals for the association of the lack of the sharing of housework (COMP) with high scores of the QMPA (SCOR) based on logistic regression models.

Especificação	Nº	RP	Método Delta 95% IC	OR	Método Woolf 95% IC
Associação					
Bruta	611	1,90	(1,34 2,69)	2,28	(1,48 3,50)
Estado civil					
Casadas	362	1,55	(1,00 2,41)	1,84	(1,03 3,27)
Solteiras	249	1,72	(0,95 3,12)	1,90	(0,94 3,83)
Ajustada		1,62	(1,13 2,32)	2,09	(1,35 3,22)
Educação					
Menor/elementar	329	1,47	(0,98 2,21)	1,72	(0,99 2,98)
Maior/elementar	282	2,32	(1,29 4,54)	2,83	(1,38 5,82)
Ajustada		1,78	(1,25 2,30)	1,86	(1,19 2,91)
Ajustada para estado civil e educação	611	1,61	(1,12 2,30)	1,83	(1,17 2,86)

RP - Razão de Prevalências; ORP - Odds Ratio da Prevalência; 95%IC - Intervalo de Confiança para 95%.
QMPA (SCOR) - Questionário de Morbidade Psiquiátrica de Adultos - escore.

Tabela 4 - Razões de Prevalência, Odds Ratio e Intervalos de Confiança para 95% para a associação entre Falta do Compartilhamento do Trabalho Doméstico (COMP) e Escores Altos do QMPA (SCOR) com o controle simultâneo de duas variáveis de confusão, baseados na análise estratificada.

Table 4 - Prevalence Ratios, Odds Ratio and 95% Confidence Intervals for the association the lack of the sharing of housework (COMP) with high scores of the QMPA (SCOR), controlling for two confounders by stratified analysis.

Variáveis		Estado civil			
		Casadas		Solteiras	
Educação		Razão	95% IC	Razão	95% IC
Menor/elementar	RP	1,30	(0,83 2,03)	1,93	(0,79 4,76)
	ORP	1,47	(0,77 2,78)	2,32	(0,77 7,02)
Maior/elementar	RP	6,47	(0,93 44,95)	1,08	(0,44 2,70)
	ORP	8,75	(1,13 68,02)	1,10	(0,39 3,05)
Ajustada p/ estado civil e educação		RP _{Aj}	1,56		(1,10 2,21)
		ORP _{Aj}	1,80		(1,14 2,86)

RP - Razão de Prevalências; ORP - Odds Ratio da Prevalência;

95%IC - Intervalo de Confiança para 95%. RP_{Aj} e ORP_{Aj} - medidas ajustadas

QMPA (SCOR) - Questionário de Morbidade Psiquiátrica de Adultos.

Tabela 5 - Razões de Prevalência, Odds Ratio e Intervalos de Confiança para 95% para a associação entre Falta do Compartilhamento do Trabalho Doméstico (COMP) e Escores Altos do QMPA (SCOR) com o controle simultâneo de duas variáveis de confusão, baseados na análise logística e Método Delta.

Table 5 - Prevalence Ratios, Odds Ratio and 95% Confidence Intervals for the association of the lack of the sharing housework (COMP) with high scores of the QMPA (SCOR), controlling for two confounders through logistic regression and the Delta Method.

Variáveis		Estado civil			
		Casadas		Solteiras	
Educação		Razão	95% IC	Razão	95% IC
Menor/elementar	RP	1,30	(0,83 2,03)	1,93	(0,79 4,76)
	ORP	1,47	(0,77 2,78)	2,32	(0,77 7,02)
Maior/elementar	RP	6,47	(0,93 44,95)	1,08	(0,44 2,70)
	ORP	8,75	(1,13 68,02)	1,10	(0,39 3,05)
Ajustada p/ estado civil e educação		RP _{Aj}	1,61		(1,12 2,31)
		ORP _{Aj}	1,83		(1,17 2,86)

RP - Razão de Prevalências; ORP - Odds Ratio da Prevalência; 95%IC - Intervalo de Confiança para 95%. RP_{Aj} e ORP_{Aj} - medidas ajustadas

QMPA (SCOR) - Questionário de Morbidade Psiquiátrica de Adultos.

CONCLUSÕES

Os fundamentos teóricos e resultados mostrados no presente trabalho apóiam o uso do método delta para estimar intervalos de confiança para razões de proporções com as matrizes de covariância geradas com a regressão logística, que produz valores aproximadamente iguais àqueles que resultam do método tabular. Razões de proporções podem, por sua vez, ser obtidas através do método convencional indicado nos livros-textos^{3, 7}, que se baseia na estimativa das probabilidades esperadas, conforme é mostrado nas equações apresentadas na Tabela 1.

Atenção deve ser dada aos limites da exploração feita no presente artigo: é analisada uma situação com uma prevalência especificada, na ausência de interação e com apenas duas variáveis de confusão, ambos categoriais e dicotômicos. É necessário, a ve-

rificação da aplicação e adequação desse método com diferentes valores de prevalência do efeito, desde que estas são conhecidas como positivamente associadas com o desvio entre OR e razões de proporções. Situações onde interações estejam presentes poderão ser abordadas através de uma análise estrato-específica. É também possível apontar que aproximações dos valores ajustados para confusão apresentem, consistentemente, afastamentos porque os pesos conferidos para a obtenção de medidas sumarizadas ajustadas modifica-se de acordo com o método empregado³. Por fim, recomenda-se o uso cuidadoso de técnicas de modelagem em análise epidemiológica, que deve se limitar a situações onde seja necessário o ajuste simultâneo de variáveis de confusão, na condição de que a análise estratificada seja inviável. A análise estratificada é mais vantajosa pela transparência e clareza dos procedimentos e, frequen-

temente, por ser de fácil interpretação, não requerendo a adoção dos muitos pressupostos como na modelagem matemática. Desse modo, os métodos apresentados no presente artigo tornam viável a estimação, com as vantagens inerentes à modelagem, da razão de prevalência e intervalos de confiança em estudos do tipo corte transversal ou coorte.

AGRADECIMENTOS

Ao Dr. Sander Greenland, da University of Califórnia (USA). Ao Dr. Bikas K. Sinha, do Indian Statistical Institute, Índia, ao Dr. Dana Loomis, University of North Carolina at Chapel Hill, USA, pelo estímulo à realização desse trabalho.

REFERÊNCIAS BIBLIOGRÁFICAS

1. AXELSON, O.; FREDERIKSON, M.; EKBERG, K. Use of the prevalence ratio v prevalence odds ratio as a measure of risk in cross-sectional studies. *Occup. Environ. Med.*, **51**:143-4, 1994.
2. BISHOP, Y.M.M.; FIENBERG, S.E.; HOLLAND, P.W. *Discrete multivariate analysis: theory and practice*. Cambridge, MA, The MIT Press, 1975.
3. CHECKOWAY, H.; PEARCE, N.; CRAWFORD-BROWN, D. J. *Research methods in occupational epidemio*. New York, Oxford University Press, 1989.
4. GREENLAND, S. Interpretation and choice of effect measures in epidemiologic analysis. *Am. J. Epidemiol.*, **125**:761-8, 1987.
5. GREENLAND, S. & ROBINS J.M. Estimation of a common effect parameter from sparse follow-up data. *Biometrics*, **41**:55-68, 1985.
6. GREENLAND, S. & THOMAS, D.C. On the need for rare disease assumption in case control in case-control studies. *Am. J. Epidemiol.*, **116**:547-53, 1982.
7. HOSMER JR., D. W. & LEMESHOW, S. *Applied logistic regression*. New York, John Wiley & Sons, 1989.
8. KLEINBAUM, D.G; KUPPER, L.L; MORGENSTERN, H. *Epidemiologic research: principles and quantitative methods*. New York, Van Nostrand Reinhold, 1982.
9. LEE, J. & CHIA, K.S. Estimation of prevalence rate ratios for cross-sectional data: an example in occupational epidemiology. *Br. J. Ind. Med.*, **50**:861-2, 1993.
10. LEE, J. Odds ratio or relative risk for cross-sectional data? *Int. J. Epidemiol.*, **23**:201-3, 1994.
11. MIETTINEN, O. Estimability and estimation in case-referent studies. *Am. J. Epidemiol.*, **103**:226-35, 1976.
12. RODRIGUES, L. & KIRKWOOD, B. R. Case-control designs in the study of common diseases: updates on the demise of the rare disease assumption and the choice of sampling scheme for controls. *Int. J. Epidemiol.*, **19**:205-13, 1990.
13. SAS. *SAS/IML Software, Version 6*. Cary, NC, SAS Institute Inc., 1990.
14. SAVITZ, D.A. Measurements, estimates and inferences in reporting epidemiologic study results. *Am. J. Epidemiol.*, **135**:223-24, 1992.
15. STROMBERG, U. Prevalence odds ratio v prevalence ratio. *Occup. Environ. Med.*, **51**:143-4, 1994.

APÊNDICE 1

Estimativa de Var[ln(\hat{r})]

Tem-se inicialmente $\ln(\hat{r}) = \ln \frac{\hat{p}_1}{\hat{p}_0} = \ln \hat{p}_1 - \ln \hat{p}_0$.

$\text{Vâr}[\ln(\hat{r})] = \text{Vâr}[\ln(\hat{p}_1)] + \text{Vâr}[\ln(\hat{p}_0)] - 2 \text{Côv}[\ln(\hat{p}_1), \ln(\hat{p}_0)]$.

Desenvolvendo-se as funções $\ln \hat{p}_1$ e $\ln \hat{p}_0$ em Séries de Taylor, obtém-se

$\text{Vâr}[\ln(\hat{p}_1)] \approx \text{Vâr}(\hat{p}_1) \times \frac{1}{\hat{p}_1^2}$;

$\text{Vâr}[\ln(\hat{p}_0)] \approx \text{Vâr}(\hat{p}_0) \times \frac{1}{\hat{p}_0^2}$;

$\text{Côv}[\ln(\hat{p}_1), \ln(\hat{p}_0)] \approx \text{Côv}(\hat{p}_1, \hat{p}_0) \times \frac{1}{\hat{p}_1} \times \frac{1}{\hat{p}_0}$.

Disto resulta,

$\text{Vâr}[\ln(\hat{r})] \approx \text{Vâr}(\hat{p}_1) \times \frac{1}{\hat{p}_1^2} + \text{Vâr}(\hat{p}_0) \times \frac{1}{\hat{p}_0^2} - 2 \text{Côv}(\hat{p}_1, \hat{p}_0) \times \frac{1}{\hat{p}_1} \times \frac{1}{\hat{p}_0}$

Do mesmo modo, desenvolvendo-se a função

$(\hat{p}_0, \hat{p}_1, \dots, \hat{p}_r)$ de (7), com $X_1 = 1$ e $X_1 = 0$, pode-se obter,

$\text{Vâr}(\hat{p}_1) \approx \sum_{i=0}^r \text{Vâr}(\hat{p}_i) \frac{1}{\hat{p}_i} + \sum_{i,j} \text{Côv}(\hat{p}_i, \hat{p}_j) \frac{1}{\hat{p}_i} \cdot \frac{1}{\hat{p}_j}$,

$\text{Vâr}(\hat{p}_0) \approx \sum_{i=0}^r \text{Vâr}(\hat{p}_i) \frac{0}{\hat{p}_i} + \sum_{i,j} \text{Côv}(\hat{p}_i, \hat{p}_j) \frac{0}{\hat{p}_i} \cdot \frac{0}{\hat{p}_j}$,

$\text{Côv}(\hat{p}_1, \hat{p}_0) \approx \sum_{i=0}^r \text{Vâr}(\hat{p}_i) \frac{1}{\hat{p}_i} \cdot \frac{0}{\hat{p}_i} +$

$+ \sum_{i,j} \text{Côv}(\hat{p}_i, \hat{p}_j) \frac{1}{\hat{p}_i} \cdot \frac{0}{\hat{p}_j}$,

onde $\hat{p}_1 = (\cdot) \Big|_{x_1=1}$; $\hat{p}_0 = (\cdot) \Big|_{x_1=0}$ e (\cdot)

está definida em (7).

Os valores das derivadas parciais são dados por:

$\frac{\partial}{\partial \hat{p}_i} = \hat{p}_1 \cdot \hat{q}_1 \cdot X_i$, onde $\hat{q}_1 = 1 - \hat{p}_1$; $\frac{\partial}{\partial \hat{p}_j} = \hat{p}_1 \cdot \hat{q}_1 \cdot X_j$, $i, j = 0, 1, 2, \dots, r$.

$\frac{\partial}{\partial \hat{p}_i} = \hat{p}_0 \cdot \hat{q}_0 \cdot X_i$, onde $\hat{q}_0 = 1 - \hat{p}_0$; $\frac{\partial}{\partial \hat{p}_j} = \hat{p}_0 \cdot \hat{q}_0 \cdot X_j$, $i, j = 0, 1, 2, \dots, r$.

Agora, sejam

$X = [X_0, X_1, X_2, \dots, X_r]$ e $\hat{p} =$

$$\hat{p} = \begin{pmatrix} \text{Vâr}(\hat{p}_0) & \text{Côv}(\hat{p}_0, \hat{p}_1) & \dots & \text{Côv}(\hat{p}_0, \hat{p}_r) \\ \text{Côv}(\hat{p}_1, \hat{p}_0) & \text{Vâr}(\hat{p}_1) & \dots & \text{Côv}(\hat{p}_1, \hat{p}_r) \\ \vdots & \vdots & \ddots & \vdots \\ \text{Côv}(\hat{p}_r, \hat{p}_0) & \text{Côv}(\hat{p}_r, \hat{p}_1) & \dots & \text{Vâr}(\hat{p}_r) \end{pmatrix}$$

Então

$\text{Vâr}(\hat{p}_1) = (\hat{p}_1 \hat{q}_1)^2 X_1 \hat{X}_1$

$\text{Vâr}(\hat{p}_0) = (\hat{p}_0 \hat{q}_0)^2 X_0 \hat{X}_0$, onde $X_1 = [1, 1, X_2, \dots, X_r]$ e $X_0 = [1, 0, X_2, \dots, X_r]$.

Portanto,

$\text{Vâr}[\ln(\hat{r})] = \hat{q}_1^2 X_1 \hat{X}_1 + \hat{q}_0^2 X_0 \hat{X}_0 - 2 \hat{q}_1 \hat{q}_0 X_1 \hat{X}_0$

Finalmente,

$\text{Vâr}[\ln(\hat{r})] = X_r \hat{X}_r$ onde $X_r = \hat{q}_1 X_1 - \hat{q}_2 X_0$ (9)

APÊNDICE 2

PROGRAMA SAS/IML (VERSÃO 6.04) PARA O CÁLCULO DAS RAZÕES DE PROPORÇÕES E INTERVALOS DE CONFIANÇA

```

*/—Este programa produz Intervalos de Confiança para a Razão de Proporções
*/—estimada a partir de um modelo de Regressão Logística, utilizando o Método
*/—Delta para o cálculo da variância dos estimadores dos parâmetros do modelo
*/—Os dados para implementação são do exemplo apresentado no texto.
*/—As palavras em itálico são comandos do SAS, enquanto que as demais
*/—deverão ser fornecidas pelo usuário. Notar que os valores 0,657 e 0,593
*/—correspondem, respectivamente, às proporções das categorias de interesse
*/—das covariáveis (confundidoras) na amostra que foram codificadas como 1.
*/—Referem-se, portanto, a proporção de casadas (ECIV=1) e ao de mulheres
*/—com educação elementar ou analfabetas (EDC=1), permitindo o ajuste da
*/—associação principal para essas covariáveis através do modelo.

options ps=64 ls=78; /* entrada de dados
data one; /* que dependerá
infile 'b:\artigo1.dat'; /* do formato ou
input comp eciv scor; /* configuração do banco

proc logistic descending outest = covmat covout; /* procedimentos para a
model scor = comp eciv edc / covb; /* estimação de parâmetros
run; /* e matrix de variâncias e
/* covariâncias do modelo
/* logístico

proc iml; /* definição das
use covmat; /* matrizes de variâncias e
read all var{intercep comp eciv edc} /* covariâncias e
into matcov /* covariâncias dos
where (_type_ = 'COV'); /* coeficientes
read all var{intercep comp eciv edc} /*
into matcoef /*
where(_type_ = 'PARMS'); /*

beta0 = matcoef[,1]; /*
beta1 = matcoef[,2]; /*
beta2 = matcoef[,3]; /*
beta3 = matcoef[,4]; /*

x1 = { 1 1 0,657 0,593 }; /* vetores de dados
x0 = { 1 0 0,657 0,593 }; /* para os grupos de
/* expostos e não expostos

p1hat = exp(beta0+beta1+beta2*0.657+ /* estimação da
beta3*0.593)/ /* prevalência nas expostas
(1+exp(1+beta0+beta1+beta2*0.657+ /*
beta3*0.593)); /*

p0hat = exp(beta0+beta2*0.657+beta3*0.593)/ /* e entre as não expostas
(1+exp(beta0+beta2*0.657 +beta3*0.593)) /*

rp = p1hat/p0hat; lnrp = log(rp) /* estimação da razão de
/* de proporções

xstar = (1-p1hat)*x1-(1-p0hat)*x0; /* estimação da variância
varlogrp = xstar*matcov*xstar'; /* da razão de proporções
stderr = sqrt (varlogrp) /* e do erro padrão

lb = exp(lnrp-1.96*stderr); /* limites de 95%
ub = exp(lnrp+1.96*stderr); /* de confiança
print pr lnrp stderr lb ub; /* impressão dos
quit; /* resultados

```