

SENSIBILIDADE E ESPECIFICIDADE DE TESTES DIAGNÓSTICOS EM DIFERENTES CENÁRIOS DE DEPENDÊNCIA

Nívea Bispo da SILVA¹
Leila Denise Alves Ferreira AMORIM²
Rosemeire Leovigildo FIACCONE²

- **RESUMO:** A avaliação de testes diagnósticos e validação de um instrumento estão centradas na estimação da sensibilidade e especificidade. Na prática, mais de dois testes diagnósticos podem ser realizados, tanto sequencialmente como simultaneamente, no mesmo indivíduo e pode ser de interesse avaliar suas performances em relação ao procedimento padrão-ouro. Uma outra situação peculiar ocorre quando um teste diagnóstico é repetido múltiplas vezes para um mesmo indivíduo. Nos últimos anos tem havido um aumento do número de pesquisas relacionadas à avaliação de testes diagnósticos envolvendo dados agrupados ou com alguma estrutura de dependência. Diversos métodos estatísticos têm sido sugeridos e avaliados como forma de corrigir o viés e a variância das estimativas da sensibilidade e especificidade em diferentes estruturas de dependência. Neste trabalho são sumarizadas e comparadas algumas das metodologias disponíveis na literatura para estimação pontual e intervalar para sensibilidade e especificidade quando existem mais de dois testes diagnósticos realizados no mesmo indivíduo, ou quando há mais de uma medida do mesmo teste para cada indivíduo. Estudos de simulação são conduzidos e uma aplicação com análise de dados sobre infecção por *Ascaris* é apresentada. Em qualquer que seja a situação de dependência dos dados, os métodos de estimação precisam levar em consideração a correlação existente entre as múltiplas medidas de um mesmo indivíduo. Definição de novos métodos ainda são necessários para lidar com o aumento da complexidade da estrutura dos dados dos estudos considerando, por exemplo, múltiplas mensurações em múltiplos testes diagnósticos.
- **PALAVRAS-CHAVE:** Sensibilidade; especificidade; dados correlacionados; múltiplos testes; múltiplas mensurações.

¹Universidade Federal de Minas Gerais - UFMG, Departamento de Estatística, CEP: 31270-901, Belo Horizonte, MG, Brasil. E-mail: nivea.bispo@gmail.com

²Universidade Federal da Bahia - UFBA, Departamento de Estatística, CEP: 40170-110, Salvador, BA, Brasil. E-mail: leiladen@ufba.br, fiacccone@ufba.br

1 Introdução

A avaliação de testes diagnósticos e validação de um instrumento estão centradas na estimação da sensibilidade e especificidade (MOONS et al., 1999, 2004; SCHUNEMANN et al. 2008), que mensuram a capacidade de um instrumento ou teste classificarem corretamente os indivíduos submetidos aos mesmos. A sensibilidade e especificidade são medidas importantes quando se deseja determinar a eficácia de um teste realizado para detectar a presença ou ausência de uma doença ou outro evento qualquer. O procedimento mais comumente usado para avaliação de um único teste diagnóstico é através de sua comparação com resultados de outro teste referenciado pela literatura como padrão-ouro (MENEZES e SANTOS, 1999). Na prática, mais de dois testes diagnósticos podem ser realizados, tanto sequencialmente como simultaneamente, no mesmo indivíduo e pode ser de interesse avaliar suas performances em relação ao procedimento padrão-ouro. Uma outra situação peculiar ocorre quando um teste diagnóstico é repetido múltiplas vezes para um mesmo indivíduo. Neste caso tem-se uma estrutura de dados agrupados, onde as observações repetidas em um mesmo indivíduo ou teste formam *clusters* (GENDERS et al, 2012). Métodos convencionais que assumem que todas as observações são independentes (estimador binomial) podem não ser válidos para estimar a variância da sensibilidade e da especificidade nestas situações, pois ignoram a correlação existente entre tais medidas, resultando em estimadores viesados para suas variâncias (GENÇ et al, 2004, GENDERS et al, 2012). A estrutura de dependência dos dados precisa ser levada em consideração sobretudo para se fazer inferência sobre as medidas diagnósticas, particularmente nestes casos, na construção dos intervalos de confiança comumente associados à sensibilidade e especificidade (STERNBERG e HADGU, 2001; GENÇ et al, 2004; YU et al, 2008; GENDERS et al, 2012).

Nos últimos anos tem havido um aumento do número de pesquisas relacionadas à avaliação de testes diagnósticos envolvendo dados agrupados em diferentes áreas do conhecimento, como por exemplo em radiologia, odontologia, imunologia e ortopedia. Um exemplo que envolve a análise da acurácia de um teste diagnóstico usando múltiplas mensurações no mesmo paciente relaciona-se ao rastreamento para localizar e detectar lesões na mamografia. Neste caso a imagem da mamografia é dividida em pequenas áreas pré-definidas para possibilitar a localização e detecção de uma possível lesão maligna. Situação similar refere-se à detecção de doença periodontal através de exames radiográficos que, em geral, são realizados em locais distintos do dente para auxiliar no diagnóstico e na definição da extensão da doença (GENDERS et al, 2012). Há também as situações onde as múltiplas mensurações em um indivíduo são decorrentes da aplicação de diferentes testes a serem usados para um mesmo diagnóstico clínico, o que requer a estimação de múltiplas sensibilidades e especificidades (LEISENRING et al., 1997). Exemplos incluem o diagnóstico de leishmaniose visceral em cães, que pode ser feita através de diferentes métodos parasitológicos, sorológicos ou moleculares (SOLCÀ et al., 2014) ou ainda diferentes critérios para definição de doença periodontal (GOMES-FILHO et al., 2007).

Métodos que ignoram a estrutura de *cluster* dos dados tendem a subestimar (ou superestimar) o erro-padrão da medida de interesse, levando a conclusões errôneas em termos de inferência estatística. Muitas vezes as observações dentro de um *cluster* são positivamente (ou negativamente) correlacionadas, podendo resultar em estimativas tendenciosas da sensibilidade e especificidade quando esta dependência é ignorada (GENÇ et. al., 2004; GENDERS et. al., 2012). Diversos métodos estatísticos têm sido sugeridos e avaliados como forma de corrigir o viés e a variância das estimativas da sensibilidade e especificidade em diferentes estruturas de dependência (Sternberg & Hadgu, 2001; McCarthy e Guo, 2005; Leon et al., 2009). Propostas incluem a definição de um estimador razão para a variância dessas medidas (McCARTHY e GUO, 2005); método baseado em verossimilhança para estimação em dados emparelhados (LEON et al, 2009), ou ainda o ajuste da variância utilizando o modelo logístico de efeitos aleatórios e as equações de estimação generalizadas (EEG) em dados de conglomerados (GENDERS et. al., 2012). Os estimadores mais comumente usados para estimação da variância da sensibilidade e da especificidade, sem inclusão de covariáveis, quando existe uma estrutura de correlação são: estimador razão, estimador correlação dentro do *cluster* e o estimador razão ponderado (AHN, 1997). Yu et al.(2008) propuseram uma metodologia que leva em consideração a correlação existente entre mais de dois testes diagnósticos realizados no mesmo indivíduo, e conseqüentemente a estimação de múltiplas sensibilidades e especificidades, enquanto Leisenring et al. (1997) propuseram o uso de modelos de regressão marginal usando EEG. As EEGs têm sido frequentemente usadas para estimação pontual e construção de intervalos de confiança para sensibilidade e especificidade com dados binários correlacionados (STERNBERG e HADGU, 2001). Em tais casos mais de um procedimento pode ser utilizado para obtenção dos intervalos de confiança: através do uso de transformação da função logito, ou através do método delta. Meier et al. (2012), por exemplo, fizeram uso das EEG's para comparar a performance de dois testes diagnósticos em pacientes com distúrbios reumatológicos.

Índices para estudar a dependência entre dois testes diagnósticos em situações onde não é possível identificar o verdadeiro status da doença em indivíduos com resposta negativa também foram propostos (BOHNING e PATILEA, 2008). Ao mesmo tempo, outros autores incorporaram variáveis latentes em modelagem estatística para avaliar a acurácia de múltiplos testes diagnósticos em situações em que não há padrão-ouro (QU et al, 1996; QU e HADGU, 1998) ou para fazer inferência onde múltiplos testes diagnósticos são considerados (DENDUKURI e JOSEPH, 2001). Uma outra abordagem é usando padrões de referência compostos pela combinação de múltiplos testes na ausência de um padrão-ouro perfeito (SCHILLER et al., 2016). A abordagem bayesiana para permitir a correlação entre dois testes diagnósticos também tem sido discutida na literatura (GEORGIADIS et al., 2003; TOVAR e ACHCAR, 2011).

O objetivo deste artigo é sumarizar e sistematizar algumas das metodologias disponíveis na literatura para estimação pontual e intervalar para sensibilidade e especificidade quando existem mais de dois testes diagnósticos realizados no mesmo

indivíduo, ou quando há mais de uma medida do mesmo teste para cada indivíduo. Maior parte da literatura referente à análise de dados correlacionados não discute estratégias de análises específicas para sensibilidade e especificidade, ou é limitada a algum tipo particular de estrutura de dependência. Com este trabalho esperamos ampliar o debate sobre a importância de se levar em consideração a correlação na realização de inferência para as medidas mais comumente usadas para avaliação de testes diagnósticos e potencialmente estimular o desenvolvimento de novos métodos estatísticos para casos omissos na literatura.

O artigo está organizado da seguinte forma: Na seção 2 as metodologias para estimação da sensibilidade e especificidade são apresentadas, sendo que na subseção 2.2 são definidos os métodos de estimação para múltiplos testes, e na subseção 2.2 apresentamos os métodos de estimação para múltiplas mensurações. A seção 3 traz as aplicações e principais resultados obtidos. A metodologia para múltiplos testes é discutida através de um estudo para avaliação do impacto de um programa de saneamento ambiental na saúde da população (subseção 3.1). Na subseção 3.2 a estimação de múltiplas mensurações é ilustrada através de estudos de simulação. Por fim, na seção 4 são apresentadas as considerações finais do artigo.

2 Material e métodos

Para as formulações a seguir, considere que se deseja estimar a sensibilidade e especificidade de J testes diagnósticos realizados em cada um dos indivíduos, e que foram obtidas informações referentes a covariáveis desses indivíduos. Considere ainda que, em algumas situações, se deseja estimar a sensibilidade e especificidade quando existem múltiplas mensurações ou testes para o mesmo indivíduo. As metodologias para cada caso com estrutura de dependência são descritas a seguir.

2.1 Métodos de estimação para múltiplos testes

Quando se trabalha com dados que possuem medidas repetidas, ou seja, valores de uma ou mais variáveis resposta, categorizadas ou contínuas, resultantes de múltiplas mensurações, podemos utilizar Modelos Lineares Generalizados - MLG (NELDER e WEDDERBURN, 1972). Liang e Zeger (1986) propuseram um método unificado para respostas contínuas e discretas baseado nas Equações de Estimação Generalizadas - EEG - para obter estimativas consistentes e assintoticamente normais para os parâmetros de um modelo de regressão, sob a especificação correta da função de ligação, tratando os parâmetros de correlação como parâmetros de perturbação.

Para definição do método de interesse, seja y_{ij} o resultado (positivo ou negativo) do j -ésimo teste diagnóstico para o i -ésimo indivíduo, de modo que $y_i^T = (y_{i1}, y_{i2}, \dots, y_{in_i})$ é o vetor com n_i respostas binárias para o i -ésimo indivíduo, e x_{ij}^T o vetor $p \times 1$ de covariáveis associadas a y_{ij} , tal que $x_i^T = (x_{i1}, x_{i2}, \dots, x_{in_i})$ é uma matriz de covariáveis com dimensão $p \times n_i$. Nesta notação o número de testes para os quais se têm resultado para cada indivíduo (n_i) pode variar. Para

estimação do vetor de parâmetros β associado a p covariáveis utiliza-se as EEG que correspondem a uma função escore quando se especifica de forma correta a matriz de variância-covariâncias V_i dos y_i . A referida matriz foi parametrizada por Liang e Zeger (1986) em termos das correlações, através da introdução de uma matriz de correlação de trabalho, denotada por $R_i(\alpha)$, com o objetivo de incorporar a correlação dentro do vetor de respostas repetidas y_{in_i} referente ao i -ésimo indivíduo.

A matriz V_i é expressa como:

$$V_i = A_i^{1/2} R_i(\alpha) A_i^{1/2} \phi,$$

em que A_i é uma matriz diagonal $n_i \times n_i$, o parâmetro α na matriz R_i representa o modelo especificado para $corr(y_{ij}, y_{jk})$, e ϕ é o parâmetro de dispersão (LIANG e ZEGGER, 1986). Portanto, a função de estimação EEG é dada pela seguinte expressão:

$$U(\beta) = \sum_{i=1}^{n_i} \frac{\partial \mu_i^T}{\partial \beta} V_i^{-1} (y_i - \mu_i) = 0, \quad (1)$$

em que μ_i é a média de y_i . No caso genérico do modelo para respostas binárias, $\mu_i = P(y_i = 1 | X_i) = \pi_i$ é a probabilidade de um resultado particular de interesse ser observado.

Para estimarmos a sensibilidade e especificidade de um teste arbitrário j , pode-se ajustar o seguinte modelo de regressão logística:

$$\log \left(\frac{\mu_{ij}}{1 - \mu_{ij}} \right) = \beta_0 + \beta_1 g_i + \beta_2 t_{i2} + \dots + \beta_J t_{iJ} + \beta_* x_{ij}^T, \quad (2)$$

em que $\mu_{ij} = P(y_{ij} = 1 | x_{ij}, g_i, t_{ij})$, g_i corresponde ao resultado do teste padrão-ouro para o i -ésimo indivíduo, t_{ij} é um indicador de que o i -ésimo indivíduo realizou o j -ésimo teste diagnóstico, $j=2, \dots, J$, sendo arbitrariamente o teste $j = 1$ aquele de referência, e x_{ij}^T contém covariáveis que podem estar associadas com a resposta.

Deste modo, a sensibilidade e a especificidade podem ser estimadas, respectivamente, através das seguintes expressões (STENBERG e HADGU, 2001):

$$P(y_{ij} = 1 | g_i = 1, x_{ij}) = \pi_{11j} = \frac{e^{\beta_0 + \beta_1 + \beta_j + \beta_* x_{ij}}}{1 + e^{\beta_0 + \beta_1 + \beta_j + \beta_* x_{ij}}} \quad e$$

$$P(y_{ij} = 0 | g_i = 0, x_{ij}) = \pi_{00j} = \frac{1}{1 + e^{\beta_0 + \beta_j + \beta_* x_{ij}}},$$

em que β_1 e β_j representam, respectivamente, o efeito do teste padrão-ouro e do j -ésimo teste diagnóstico, $j=2, 3, \dots, J$. β_* é um vetor $p \times 1$ de coeficientes para x_{ij} . Para o teste de referência, $j = 1$, tem-se que as correspondentes sensibilidade e especificidade são definidas por:

$$\pi_{11j} = \frac{e^{\beta_0 + \beta_1 + \beta_* x_{ij}}}{1 + e^{\beta_0 + \beta_1 + \beta_* x_{ij}}} \quad \text{e} \quad \pi_{00j} = \frac{1}{1 + e^{\beta_0 + \beta_* x_{ij}}}.$$

O cálculo dos intervalos de confiança para as sensibilidades e especificidades referentes ao j -ésimo teste diagnóstico pode ser feito utilizando-se o método da transformação a partir da função logito, ou ainda através do método delta. O modelo (2) também pode incluir um efeito de interação entre g_i e t_{ij} .

Para o método da transformação, considere:

$$h_{1j}(\beta) = \text{logito}(\pi_{11j}) = \log\left(\frac{\pi_{11j}}{1 - \pi_{11j}}\right) \quad \text{e} \quad h_{0j}(\beta) = \text{logito}(\pi_{00j}) = \log\left(\frac{\pi_{00j}}{1 - \pi_{00j}}\right),$$

em que h_{1j} e h_{0j} representam, respectivamente, os logitos da sensibilidade e especificidade para o j -ésimo teste diagnóstico.

Os intervalos de $(1 - \alpha)\%$ de confiança para sensibilidade e especificidade, baseados na normalidade assintótica do vetor de parâmetros definido em (2), podem ser expressos por:

$$\frac{e^{h_{kj}(\hat{\beta}) \pm z_{\alpha/2} \sqrt{\widehat{\text{Var}}(h_{kj}(\hat{\beta}))}}}{1 + e^{h_{kj}(\hat{\beta}) \pm z_{\alpha/2} \sqrt{\widehat{\text{Var}}(h_{kj}(\hat{\beta}))}}}, \quad k = 0, 1.$$

em que $\text{Var}(h_{kj}(\hat{\beta}))$ denota a variância da combinação linear dos estimadores dos coeficientes da regressão para o j -ésimo teste, sendo obtida a partir da matriz de variância-covariância robusta estimada pelas EEG's.

O método delta, por sua vez, considera aproximações de primeira ordem da variância da sensibilidade e especificidade, tal que:

$$\text{Var}_{\Delta}(\hat{\pi}_{kkj})(\hat{\beta}) \simeq \Delta_{kj} V_{EEG} \Delta_{kj}^T, \quad k = 0, 1 \quad \text{e} \quad j = 1, 2, \dots, J. \quad (3)$$

em que V_{EEG} é a variância robusta obtida pelas EEG's. Além disso:

$$\Delta_{0j} = \left(\frac{\partial \pi_{00j}}{\partial \beta_0}, \dots, \frac{\partial \pi_{00j}}{\partial \beta_{2J+p-1}} \right) \quad \text{e} \quad \Delta_{1j} = \left(\frac{\partial \pi_{11j}}{\partial \beta_0}, \dots, \frac{\partial \pi_{11j}}{\partial \beta_{2J+p-1}} \right).$$

Assim, os intervalos de $(1 - \alpha)\%$ de confiança podem ser calculados assumindo normalidade assintótica para as sensibilidades e especificidades estimadas, tal que:

$$\left[\hat{\pi}_{kkj} - z_{\alpha/2} \sqrt{\widehat{\text{Var}}_{\Delta}(\hat{\pi}_{kkj}(\hat{\beta}))}; \hat{\pi}_{kkj} + z_{\alpha/2} \sqrt{\widehat{\text{Var}}_{\Delta}(\hat{\pi}_{kkj}(\hat{\beta}))} \right],$$

$k = 0, 1 \quad \text{e} \quad j = 1, 2, \dots, J.$

Outra opção é a obtenção dos intervalos de confiança através de *bootstrap* pelo método dos percentis (STENBERG e HAGDU, 2001), mas este procedimento não é sumarizado aqui.

2.2 Métodos de estimação para múltiplas mensurações

Os estimadores mais comumente usados para estimação da sensibilidade e especificidade, e suas respectivas variâncias, em situações em que existem múltiplas mensurações do mesmo teste diagnóstico para o mesmo indivíduo são descritos a seguir. Para isto, considere a notação conforme representada na Tabela 1, onde y_i^{vp} (ou y_i^{vn}) denota o número de resultados verdadeiros-positivos (ou verdadeiros-negativos) para o i -ésimo indivíduo, e n_i^d (ou $n_i^{\bar{d}}$) o número de locais infectados/doentes (ou livres da infecção/doença) no i -ésimo indivíduo segundo o método padrão-ouro.

Tabela 1 - Representação dos dados com estrutura agregada

indivíduo	y_i^{vp}	y_i^{vn}	n_i^d	$n_i^{\bar{d}}$
1	y_1^{vp}	y_1^{vn}	n_1^d	$n_1^{\bar{d}}$
2	y_2^{vp}	y_2^{vn}	n_2^d	$n_2^{\bar{d}}$
\vdots	\vdots	\vdots	\vdots	\vdots
N	y_N^{vp}	y_N^{vn}	n_N^d	$n_N^{\bar{d}}$

$$\text{onde } \sum_{i=1}^N y_i^{vp} = y^{vp}, \sum_{i=1}^N y_i^{vn} = y^{vn}, \sum_{i=1}^N n_i^d = n^d \text{ e } \sum_{i=1}^N n_i^{\bar{d}} = n^{\bar{d}}.$$

Estimador binomial

Quando há independência das observações, o estimador binomial é o método mais comumente utilizado para obtenção das estimativas da sensibilidade e especificidade e de suas respectivas variâncias. A sensibilidade de um teste refere-se, em qualquer um das situações, à capacidade do mesmo reconhecer os verdadeiros-positivos, e a especificidade à capacidade de distinguir os verdadeiros-negativos quando comparados a um procedimento padrão-ouro. Os estimadores binomiais para a sensibilidade e especificidade podem ser obtidos, respectivamente, a partir das seguintes expressões:

$$\hat{S} = \frac{\sum_{i=1}^N y_i^{vp}}{\sum_{i=1}^N n_i^d} = \frac{y^{vp}}{n^d} \quad \text{e} \quad \hat{E} = \frac{\sum_{i=1}^N y_i^{vn}}{\sum_{i=1}^N n_i^{\bar{d}}} = \frac{y^{vn}}{n^{\bar{d}}}.$$

Esses estimadores pontuais são os mesmos para todos os métodos descritos nesta seção, onde se consideram os métodos para situações com múltiplas mensurações de um mesmo teste diagnóstico, sem incorporação de covariáveis na análise. A diferença entre os métodos consiste na definição da variância para essas medidas.

Os estimadores binomiais da variância para a sensibilidade e a especificidade, respectivamente, são:

$$Var(\hat{S}) = \frac{\hat{S}(1 - \hat{S})}{n^d} \quad \text{e} \quad Var(\hat{E}) = \frac{\hat{E}(1 - \hat{E})}{n^{\bar{d}}}. \quad (4)$$

Estimador razão

Este método foi proposto por Rao e Scott (1992) e atribui pesos iguais para os locais (*clusters*), obtendo-se as mesmas estimativas do estimador binomial (JUNG e AHN, 2000). Sob a suposição de que o número de clusters (N - número total de indivíduos) é grande, e que o tamanho de cada cluster é constante, a variância dos estimadores da sensibilidade e especificidade é, respectivamente, dada por:

$$Var_{ER}(\hat{S}) = \frac{N \sum_{i=1}^N (y_i^{vp} - n_i^d \hat{S})^2}{(N-1)(n^d)^2} \quad \text{e} \quad Var_{ER}(\hat{E}) = \frac{N \sum_{i=1}^N (y_i^{vn} - n_i^{\bar{d}} \hat{E})^2}{(N-1)(n^{\bar{d}})^2}. \quad (5)$$

Segundo Lavange et al. (1994), o estimador razão é um dos métodos mais comumente utilizados no contexto de inquéritos com estrutura de amostragem complexa, onde o interesse é estimar taxas ou proporções para subgrupos populacionais definidos pela classificação cruzada de variáveis explanatórias, e também em situações onde existem dados faltantes. Este método é válido desde que amostras de tamanho suficientemente grandes estejam disponíveis para assegurar a consistência das estimativas obtidas.

Estimador correlação dentro do cluster

Nesse método a variância é calculada levando-se em consideração a correlação entre as observações dentro de cada *cluster* (DONNER e KLAR, 1993), que é incorporada através do estimador ANOVA para o coeficiente de correlação intra-classe (JUNG et al., 2000). Esse estimador é uma versão corrigida do estimador binomial. Denotaremos o estimador do coeficiente de correlação dentro do *cluster* por $\hat{\rho}$. Os estimadores da variância para a sensibilidade e a especificidade são definidos, respectivamente, por:

$$Var_{EC}(\hat{S}) = \frac{B_{sen} \hat{S}(1 - \hat{S})}{n^d}, \quad (6)$$

em que

$$B_{sen} = \frac{\sum_{i=1}^N n_i^d L_{n_i^d} [1 + (n_i^d - 1) \hat{\rho}_{sen}]}{\sum_{i=1}^N n_i^d L_{n_i^d}} \quad \text{e} \quad \hat{\rho}_{sen} = 1 - \frac{\sum_{i=1}^N \frac{y_i^{vp} (n_i^d - y_i^{vp})}{n_i^d}}{(n^d - N) \hat{S}(1 - \hat{S})}$$

e

$$\text{Var}_{EC}(\hat{E}) = \frac{B_{esp}\hat{E}(1-\hat{E})}{n^{\bar{d}}}, \quad (7)$$

em que

$$B_{esp} = \frac{\sum_{i=1}^N n_i^{\bar{d}} L_{n_i^{\bar{d}}} [1 + (n_i^{\bar{d}} - 1)\hat{\rho}_{esp}]}{\sum_{i=1}^N n_i^{\bar{d}} L_{n_i^{\bar{d}}}} \quad \text{e} \quad \hat{\rho}_{esp} = 1 - \frac{\sum_{i=1}^N \frac{y_i^{vn}(n_i^{\bar{d}} - y_i^{vn})}{n_i^{\bar{d}}}}{(n^{\bar{d}} - N)\hat{E}(1-\hat{E})}$$

com $L_{n_i^d}$ e $L_{n_i^{\bar{d}}}$ denotando, respectivamente, o número de observações com n_i^d locais doentes e $n_i^{\bar{d}}$ locais livres da doença. $\hat{\rho}_{sen}$ e $\hat{\rho}_{esp}$ representam, respectivamente, o coeficiente de correlação dentro do *cluster* para a sensibilidade e especificidade.

Estimador razão ponderado

O estimador correlação dentro do cluster considera que um paciente que possui n_i^d locais doentes fornece menos informações que N pacientes com um local cada um. O estimador razão fornece o mesmo peso para ambos os casos, independentemente do número de locais que os pacientes possuem (Rao e Scott, 1992). Para resolver tal problema, Lee e Dubin (1994) propuseram usar o estimador razão ponderado através da introdução do peso w_i . Assim, se denotarmos $w_i^s = \frac{1}{n_i^d}$ ou $w_i^e = \frac{1}{n_i^{\bar{d}}}$ como sendo o peso para o *cluster* i , os estimadores para a sensibilidade e especificidade podem ser obtidos, respectivamente, como sendo a média simples de taxas positivas/negativas de cada *cluster*. Ou seja,

$$\hat{S}_P = \frac{1}{N} \sum_{i=1}^N w_i^s n_i^d \hat{S}_i \quad \implies \quad \hat{S}_P = \frac{\sum_{i=1}^N \hat{S}_i}{N} \quad \text{e}$$

$$\hat{E}_P = \frac{1}{N} \sum_{i=1}^N w_i^e n_i^{\bar{d}} \hat{E}_i \quad \implies \quad \hat{E}_P = \frac{\sum_{i=1}^N \hat{E}_i}{N}. \quad (8)$$

As variâncias da sensibilidade e especificidade neste método podem ser computadas como:

$$\text{Var}(\hat{S}_P) = \frac{\sum_{i=1}^N (\hat{S}_i - \hat{S}_P)^2}{N(N-1)} \quad \text{e} \quad \text{Var}(\hat{E}_P) = \frac{\sum_{i=1}^N (\hat{E}_i - \hat{E}_P)^2}{N(N-1)} \quad (9)$$

3 Aplicações e resultados

3.1 Múltiplos testes

Para implementação dos métodos discutidos na seção 2.1, são utilizados dados de 247 crianças participantes de um estudo para avaliação do impacto de um programa de saneamento ambiental na saúde da população no município de Salvador, estado da Bahia (Estudo Bahia Azul) (BARRETO et al, 2006). Diferentes métodos diagnósticos para a ocorrência de *Ascaris lumbricoides* nas crianças desse estudo foram utilizados: contagem do número de ovos de *Ascaris* em exames de fezes, níveis de IgE específica para *Ascaris* e quantificação de anticorpos IgG4, ambos obtidos através de exame de sangue. A relação entre esses marcadores e o resultado do exame de fezes foi avaliada pelo cálculo da sensibilidade e especificidade para as variáveis IgE específica e IgG4, usando a contagem do número de ovos como padrão-ouro.

A positividade para *Ascaris* foi definida quando o número de ovos foi igual ou superior a um ou quando o nível de IgE específica no sangue foi maior ou igual a 0,35 kU/I. Na literatura não há um ponto de corte para IgG4 como marcador de *Ascaris*, logo, esse foi definido através do uso de curvas ROC, tendo como padrão-ouro o número de ovos de *Ascaris*. Assim, com base nos resultados dessa análise definiu-se a positividade para *Ascaris* para valores de IgG4 superiores a 0,0039 kU/I. Do total de crianças apenas 19,4% apresentaram ovos de *Ascaris*, sendo que níveis de IgE específica no sangue maiores ou iguais a 0,35 kU/ foram encontrados em 64,3% das crianças. A estimação e construção de intervalos de confiança para a sensibilidade e a especificidade foram realizadas através do uso de EEG.

No procedimento de estimação das sensibilidades e especificidades usando EEG, y_{ij} refere-se ao resultado do j -ésimo teste diagnóstico (positivo/negativo) para o i -ésimo indivíduo, sendo modelado por:

$$\log\left(\frac{\mu_{ij}}{1 - \mu_{ij}}\right) = \beta_0 + \beta_1 g_i + \beta_2 t_{i2} + \beta_3 g_i t_{i2}$$

em que g_i corresponde ao resultado do teste padrão-ouro, e t_{i2} ao indicador de que o i -ésimo indivíduo obteve informação sobre IgG4, sendo o marcador IgE o teste de referência.

As estimativas dos β 's, bem como a matriz de variância-covariância estimada, foram:

$$\hat{\beta} = (-0,80, \quad 1,06, \quad 0,77, \quad -0,86)$$

$$\text{Var}(\hat{\beta}_{EEG}) = \begin{pmatrix} 0,02 & -0,02 & -0,02 & 0,02 \\ -0,02 & 0,11 & 0,02 & -0,09 \\ -0,02 & 0,02 & 0,03 & -0,03 \\ 0,02 & -0,09 & -0,03 & 0,18 \end{pmatrix}$$

A sensibilidade e especificidade estimadas para o IgE foram, respectivamente, 56,5% e 69,0%, enquanto que a sensibilidade e especificidade estimadas para IgG4 foram, respectivamente, 53,7% e 50,6%.

Para construção dos intervalos de confiança usando o método da transformação e o método delta, os seguintes resultados foram obtidos, considerando $j = 1$ como sendo o marcador IgE, e $j = 2$, o marcador IgG4:

$$h_{01} = 0,800; \quad h_{02} = 0,024; \quad h_{11} = 0,262; \quad h_{12} = 0,148$$

$$\hat{\Delta}_{01} = (-0,20; 0,00; 0,00; 0,00); \quad \hat{\Delta}_{11} = (0,25; 0,25; 0,00; 0,00), \quad k = 0,1; j = 1$$

$$\hat{\Delta}_{02} = (-0,25; 0,00; -0,25; 0,00); \quad \hat{\Delta}_{12} = (0,25; 0,25; 0,25; 0,25), \quad k = 0,1; j = 2$$

Os intervalos de confiança para sensibilidade e especificidade, considerando os dois métodos encontram-se na Tabela 2. Os intervalos obtidos pelos dois métodos foram muito próximos, com amplitude um pouco reduzida naqueles resultantes do uso do método da transformação. Os resultados obtidos corroboram com achados de outros autores (STERNBERG e HADGU, 2001) sobre a equivalência dos métodos para obtenção de intervalos de confiança pelas EEG's.

Tabela 2 - Intervalos de 95% de confiança para sensibilidade e especificidade para múltiplos testes para diagnóstico de *Ascaris lumbricoides*

Medidas	Método Delta		Método da Transformação	
	IgE	IgG4	IgE	IgG4
Sensibilidade	(0,42; 0,71)	(0,39; 0,68)	(0,42; 0,70)	(0,39; 0,67)
Especificidade	(0,63; 0,75)	(0,43; 0,58)	(0,62; 0,75)	(0,43; 0,58)

3.2 Múltiplas mensurações

Para os métodos discutidos na seção 2.2, foram gerados dados para ilustrar os diferentes cenários que comumente são encontrados em estudos envolvendo métodos diagnósticos com dados provenientes de múltiplos locais/múltiplas mensurações no mesmo indivíduo. Alguns dos cenários podem representar, por exemplo, vários dentes ou várias glândulas do mesmo indivíduo.

O número de verdadeiro-positivos e negativos foram gerados a partir da distribuição beta-binomial (BB), cuja probabilidade de sucesso não é constante, mas gerada a partir dos parâmetros da distribuição beta. Assim, $y_i \sim BB(n, N, \mu, \rho)$, onde y_i representa o número de verdadeiros-positivos e/ou negativos; n o número de observações; N o número de locais (sites); μ a probabilidade de sucesso, e ρ o coeficiente de correlação adotado entre os diferentes locais do mesmo indivíduo. O número de locais infectados ou livres da doença/sintoma foram gerados a partir da distribuição uniforme discreta, considerando-se que tais locais poderiam ser fixos

ou variáveis. Assim, foram geradas $n=30$ observações, considerando $N=10$ e $N=6$ locais (fixos) e o número de locais variando de 1 a 10. Os valores adotados para a sensibilidade e especificidade foram, respectivamente, 0,90 e 0,85. Além disso, foram testados dois diferentes valores para o coeficiente de correlação ρ : 0,1 e 0,7.

As Tabelas 3 e 4 apresentam, respectivamente, os intervalos de confiança para sensibilidade e especificidade quando o número de locais é fixo e igual a 10, e quando os mesmos variam de 1 a 10. Resultados similares foram obtidos quando o número de locais é fixo e igual a 6, por isso foram omitidos.

Tabela 3 - Intervalos de 95% de confiança para sensibilidade com o número de locais fixo e variando para diferentes métodos de estimação e níveis de dependência

Métodos	$\rho = 0,1$			$\rho = 0,7$		
	\hat{S}	ep	IC	\hat{S}	ep	IC
Número de locais fixo						
Binomial	0,93	0,015	(0,90; 0,96)	0,90	0,018	(0,86; 0,93)
Razão	0,93	0,022	(0,89; 0,97)	0,90	0,052	(0,80; 1,00)
Razão Ponderado	0,93	0,022	(0,89; 0,97)	0,90	0,052	(0,80; 1,00)
Correlação dentro do cluster	0,93	0,022	(0,89; 0,97)	0,90	0,052	(0,80; 1,00)
Número de locais variando						
Binomial	0,85	0,029	(0,79; 0,91)	0,87	0,027	(0,82; 0,92)
Razão	0,85	0,030	(0,79; 0,91)	0,87	0,060	(0,75; 0,99)
Razão Ponderado	0,84	0,048	(0,75; 0,93)	0,88	0,050	(0,78; 0,98)
Correlação dentro do cluster	0,85	0,034	(0,78; 0,92)	0,87	0,052	(0,76; 0,98)

Ao comparar, por exemplo, as estimativas da sensibilidade (e seus respectivos erros-padrão) quando o número de locais é fixo com o mesmo variando, observou-se que houve uma redução nas estimativas em qualquer dos níveis de dependência considerados. Além disso, a estimação do erro-padrão para a sensibilidade sofre impacto do método utilizado quando o número de locais varia.

Verificou-se que para $\rho=0,1$, por exemplo, existe pouca diferença no erro-padrão estimado pelo método binomial (0,029) e outros métodos (por exemplo, erro-padrão obtido pelo método da razão=0,030). O método que apresentou maior diferença na estimação do erro-padrão nessa situação foi o estimador razão ponderado (erro-padrão=0,048). Quando $\rho=0,7$ os resultados obtidos pelo estimador razão ponderado e correlação dentro do cluster são muito parecidos (erro-padrão=0,050 e 0,052, respectivamente, com número de locais variando).

Já na estimação da especificidade não houve impacto do método nas estimativas dos erros-padrão quando os níveis de dependência foram 0,1 e 0,7 com número de locais fixos, com exceção do método binomial. Contudo, alguma variação no erro-padrão é observada quando o número de locais varia, com método da razão apresentando a maior estimativa (erro-padrão=0,070).

Todas as análises foram realizadas no *software* R (R CORE TEAM, 2014). Para implementação do EEG na análise de dados com múltiplos testes usou-se a função `gee` do pacote `gee` do R. No apêndice encontra-se a sintaxe usada para geração dos dados do estudo de simulação.

Tabela 4 - Intervalos de 95% de confiança para especificidade com o número de locais fixo e variando para diferentes métodos de estimação e níveis de dependência

Métodos	$\rho = 0,1$			$\rho = 0,7$		
	\bar{E}	ep	IC	\bar{E}	ep	IC
Número de locais fixo						
Binomial	0,86	0,020	(0,82; 0,90)	0,79	0,023	(0,74; 0,84)
Razão	0,86	0,026	(0,80; 0,91)	0,79	0,067	(0,66; 0,92)
Razão Ponderado	0,86	0,026	(0,80; 0,91)	0,79	0,067	(0,66; 0,92)
Correlação dentro do cluster	0,86	0,025	(0,80; 0,91)	0,79	0,066	(0,66; 0,92)
Número de locais variando						
Binomial	0,89	0,025	(0,84; 0,94)	0,81	0,032	(0,75; 0,87)
Razão	0,89	0,027	(0,84; 0,94)	0,81	0,077	(0,66; 0,96)
Razão Ponderado	0,91	0,026	(0,86; 0,96)	0,85	0,059	(0,74; 0,96)
Correlação dentro do cluster	0,89	0,025	(0,84; 0,94)	0,81	0,070	(0,67; 0,95)

4 Considerações finais

Tem havido um aumento no uso da estimação intervalar para sensibilidade e especificidade em anos recentes, bem como crescente complexidade dos desenhos de estudo utilizados. Com isto, amplia-se a necessidade de discussão sobre métodos estatísticos apropriados para lidar com diferentes fontes de dependência nos dados. Neste trabalho foram discutidos métodos para estimação usando intervalos de confiança para sensibilidade e especificidade em duas situações de complexidade: (1) realização de múltiplos testes diagnósticos simultâneos; (2) realização de um mesmo teste diagnóstico em múltiplos locais/múltiplas mensurações de mesmo indivíduo. Em ambos os casos, há necessidade de utilização de estimador que leve em consideração a correlação entre as correspondentes múltiplas medidas.

Pôde-se verificar na construção dos intervalos de confiança para sensibilidade e especificidade quando múltiplos testes diagnósticos são realizados simultaneamente, que os dois métodos considerados foram muito próximos, com amplitude um pouco reduzida naqueles resultantes do uso do método da transformação. Em situações em que um mesmo teste diagnóstico é realizado em múltiplos locais do mesmo indivíduo, a escolha do método para estimação intervalar da sensibilidade e especificidade parece ser mais importante quando o número de locais varia. Independente do cenário é importante considerar métodos de estimação que levem em consideração a correlação existente entre as múltiplas medidas de um mesmo indivíduo. Os resultados obtidos corroboram com achados da literatura que recomenda o uso das EEG's para amostras de tamanho superior a 30. Os demais métodos são recomendados quando $N \leq 20$.

Pela simplicidade na implementação, quando múltiplos testes diagnósticos são realizados simultaneamente, o método da transformação é recomendado para estimação intervalar da sensibilidade e especificidade. Já nas situações onde um mesmo teste diagnóstico é realizado em múltiplos locais do mesmo indivíduo, a avaliação do comportamento dos métodos ainda precisa ser realizada em outros

cenários, considerando-se, por exemplo, amostras maiores que 100 e um número menor de locais por paciente. Outras possibilidades de análise incluem trabalhar com LCA (*Latent Classes Analysis*, em inglês), sobretudo para dar conta de situações em que não existe procedimento padrão-ouro (SOLCÀ et al, 2014) ou ainda o uso de LTA (*Latent Transition Analysis*, em inglês) para medidas repetidas ao longo do tempo (COLLINS e LANZA, 2010). O uso de modelos de efeitos mistos também têm sido discutido na literatura (LEON et al, 2007) para incorporar outras estruturas de dependência dos dados. Na mesma direção dos procedimentos baseados em modelos com a especificação da distribuição conjunta dos dados binomiais correlacionados, tem surgido na literatura abordagens utilizando cópulas com aplicação na área de Bioestatística (YIN e YUAN, 2009; KOLEV e PAIVA, 2009), não esquecendo, entretanto, da importância da interpretabilidade do parâmetro de correlação na referida abordagem. Além disso, usando a abordagem bayesiana, índices de dependência entre testes diagnósticos têm sido modelados via diferentes estruturas de cópula (TOVAR e ACHCAR, 2011; TOVAR e ACHCAR, 2013). Novos métodos ainda podem ser propostos para lidar com o aumento da complexidade da estrutura dos dados, como por exemplo, considerar múltiplas mensurações em múltiplos testes diagnósticos.

SILVA, N. B.; AMORIM, L. D. A. F.; FIACCONE, R. L. Sensitivity and specificity of diagnostic tests in different dependent scenarios. *Rev.Bras. Biom.*, São Paulo, v.34, n.3, p.489-506, 2016.

- **ABSTRACT:** *Evaluation of diagnostic tests and validation of an instrument have focused on estimation of sensitivity and specificity. In practice, more than two diagnostic tests may be performed either sequentially or simultaneously in the same subject and it may be of interest to evaluate their performance in relation to a gold standard procedure. Another peculiar situation occurs when a diagnostic test is repeated multiple times to the same individual. In recent years there has been an increase of research related to the evaluation of diagnostic tests involving grouped data or other type of data dependency. Several statistical methods have been suggested and evaluated in order to correct the bias and the variance estimates of sensitivity and specificity in different dependency structures. In this work we summarized and compared some of the methods available in the literature for point and interval estimation for sensitivity and specificity when there are more than two diagnostic tests performed in the same individual, or when more than a measure of the same test is available for each individual. Simulation studies are conducted and a data analysis related to Ascaris infection is shown. In any data dependency scenario, the estimation methods need to take into account the correlation between multiple measurements of the same individual. Definition of new methods are needed to cope with the increased complexity of the data structure of studies considering, for example, multiple measurements across multiple diagnostic tests.*
- **KEYWORDS:** *Sensibility; specificity; correlated data; multiple tests; multiple measurements.*

Referências

- AHN, C. Statistical methods for the estimation of sensitivity and specificity of site-specific diagnostic tests. *Journal of Periodontal Research*, v.32, p.351-354, 1997.
- BARRETO M. L. et al. *Evaluation of changes of sanitary conditions in sentinel areas*. In: Evaluation of the Epidemiologic Impact of the Program of Environmental Sanitation of Todos os Santos Bay (Blue Bay Program). Salvador, Brasil: Federal University of Bahia, p.232-320, 2006.
- BÖHNING, D.; PATILEA, V. A Capture-Recapture Approach for Screening Using Two Diagnostic Tests With Availability of Disease Status for the Test Positives Only. *Journal of the American Statistical Association*, v.103, p.212-221, 2008.
- COLLINS, L. M.; LANZA, S. T. *Latent Class and Latent Transition Analysis with Applications in the Social, Behavioral, and Health Sciences*. New York: John Wiley Sons, 2010.
- DENDUKURI, N.; JOSEPH, L. Bayesian Approaches to Modeling the Conditional Dependence Between Multiple Diagnostic Tests. *Biometrics*, v.57, p.158-167, 2001.
- DONNER, A.; KLAR, N. Confidence interval construction for effect measures arising from cluster randomization trials. *Journal of Clinical Epidemiology*, v.46, p.123-131, 1993.
- FREEMAN, D.; FREEMAN, J. L.; BROCAR, D.; KOCH, G. G. Strategies in the multivariate analysis of data from complex surveys II. An application to the United States National Health Interview Survey. *International Statistical Review*, v.44, p.317-330, 1976.
- GEORGIADIS, M. P.; JOHNSON, W. O.; GARDNER, I. A.; SINGH, R. Correlation adjusted estimation of sensitivity and specificity of two diagnostic tests. *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, v.52, p. 63-76, 2003.
- GENÇ, Y.; GÖKMEN, D.; TÜCCAR, E.; YAGMURLU, B. Estimation of sensitivity and specificity for clustered data. *Turkish Journal of Medical Sciences*, v.35, p.21-24, 2004.
- GENDERS, T. S. S.; SPRONK, S.; STIJNEN, T.; STEYERBERG, E. W.; LESAFFRE, E.; HUNINK, M. M. G. Methods for Calculating Sensitivity and Specificity of Clustered Data: A Tutorial. *Radiology*, v.265, p.910-916, 2012.
- GOMES-FILHO, I. S.; CRUZ, S. S.; REZENDE, E. J. C.; SANTOS, C. A. S. T.; SOLEDADE, K. R.; MAGALHÃES, M. A.; TRINDADE, S. C.; VIANNA M. I. P., PASSOS, J.; CERQUEIRA, E. M. M. Exposure measurement in the association between periodontal disease and prematurity/low birth weight. *Journal of Clinical Periodontology*, v.34, p.957-963, 2012.
- JUNG, S.; AHN, C. Estimation of response probability in correlated binary data: A new approach. *Drug Information Journal*, v.34, p.599-604, 2000.

- LAVANGE, L. M.; KEYS, L. L.; KOCH, G. G.; MARGOLIS, P. A. Application of sample survey methods for modelling ratios to incidence densities. *Statistics in Medicine*, v.13, p.343-355, 1994.
- LEISENRING, W.; PEPE, M. S.; LONGTON, G. A marginal regression modelling framework for evaluating medical diagnostic tests. *Statistics in Medicine*, v.16, p.1263-1281, 1997.
- LEE, E.; DUBIN, N. Estimation and sample size considerations for clustered binary responses. *Statistics in Medicine*, v.13, p.1241-1252, 1994.
- LEON, A. R.; GUO, O. M.; RUDNISKY, C. J.; SINGH, G. A likelihood approach to estimating sensitivity and specificity for binocular data: Application in ophthalmology. *Statistics in Medicine*, v.26, p.3300-3314, 2007.
- LIANG, K. Y.; ZEGER, S. Longitudinal data analysis using generalized linear models. *Biometrika*, v.73, p.13-22, 1986.
- MENEZES, A. B.; SANTOS, I. S. Curso de epidemiologia básica para pneumologistas: 4ª parte Epidemiologia clínica. *Jornal de Pneumologia*, v.25, p.321-326, 1999.
- MOONS, K. G.; VAN, E. A.; MICHEL, B. C.; BULLER, H. R.; HABBEMA, J. D.; GROBBEE, D. E. Redundancy of single diagnostic test evaluation. *Epidemiology*, v.10, p.276-281, 1999.
- MOONS, K. G.; GROBBEE, D. E. Diagnostic studies as multivariable, prediction research. *Journal of epidemiology and community health*, v.56, p.337-338, 2002.
- NELDER, J. A.; WEDDERBURN, R. W. M. Generalized Linear Models. *Journal of the Royal Statistical Society, Series A*, v.135, p.370-384, 1972.
- QU, Y.; TAN, M.; KUTNER, M. H. Random Effects Models in Latent Class Analysis for Evaluating Accuracy of Diagnostic Tests. *Biometrics*, v.52, p.797-810, 1996.
- QU, Y.; HADGU, A. A Model for Evaluating Sensitivity and Specificity for Correlated Diagnostic Tests in Efficacy Studies with an Imperfect Reference Test. *Journal of the American Statistical Association*, v.93, p.920-928, 1998.
- R CORE TEAM. *R: A Language and Environment for Statistical Computing*. <http://www.Rproject.org/>, Vienna-Austria, 2014.
- RAO, J. N. K.; SCOTT, A. J. A simple method for the analysis of clustered binary data. *Biometrics*, v. 48, p. 577-585, 1992.
- SCHILLER, I.; VAN SMEDEN, M.; HADGU, A.; LIBMAN, M.; REITSMA, J. B.; DENDUKURI, N. Bias due to composite reference standards in diagnostic accuracy studies. *Statistics in Medicine*, v.35, p.1454-1470, 2016.
- SCHÜNEMANN, H. J.; OXMAN, A. D.; BROZEK, J. GRADE: Grading quality of evidence and strength of recommendations for diagnostic tests and strategies. *British Medical Journal*, v.336, p.1106-1110, 2008.

- SOLCÁ, M. S. et al. Evaluating the Accuracy of Molecular Diagnostic Testing for Canine Visceral Leishmaniasis using Latent Class Analysis. *Plos One*, v.9, 2014.
- STERNBERG, M. R.; HADGU, A. A. GEE approach to estimating sensitivity and specificity and coverage properties of the confidence intervals. *Statistics in Medicine*, v.20, p.1529-1539, 2001.
- TOVAR, J. R.; ACHCAR, J. A. Indexes to Measure Dependence between Clinical Diagnostic Tests: A Comparative Study. *Revista Colombiana de Estatística*, v.34, p.433-540, 2011.
- TOVAR, J. R.; ACHCAR, J. A. Dependence between three diagnostic tests in presence of verification bias: A copula function approach. *Revista Brasileira de Biometria*, v.29, p.74-90, 2011.
- TOVAR, J. R.; ACHCAR, J. A. Dependence Between Two Diagnostic Tests with Copula Function Approach: A Simulation Study. *Communications in Statistics - Simulation and Computation*, v.42, p.454-475, 2013.
- YIN, G.; YUAN, Y. Bayesian dose finding in oncology for drug combinations by copula regression. *Applied Statistics*, v.58, p.211-224, 2009.
- YU, Q.; TANG, W.; MA, Y.; GAMBLE, S. A.; TU, X. M. Comparing multiple sensitivities and specificities with different diagnostic criteria: Applications to sexual abuse and sexual health research. *Computational Statistics & Data Analysis*, v.53, p.27-37, 2008.
- KOCH, G.; FREEMAN, D. H.; FREEMAN, J. L. Strategies in the multivariate analysis of data from complex surveys. *International Statistical Review*, v.43, p.59-78, 1975.
- KOLEV, N.; PAIVA, D. Copula based regression models: A survey. *Journal of Statistical Planning and Inference*, v.139, p.3847-3856, 2009.

Recebido em 26.02.2015.

Aprovado após revisão em 27.04.2016.

Anexo: Aspectos Computacionais do Estudo de Simulação

Nesta seção apresentamos a sintaxe para geração dos dados usados no estudo de simulação definido na seção 3.2.

Geração dos dados simulados:

<i>library(VGAM)</i>	Pacote para gerar valores da Beta-binomial
<i>m=5</i>	(número de locais infectados ou livres da doença)
<i>n=100</i>	(número de indivíduos)
<i>mu1=0.90</i>	(média da distribuição beta para os verdadeiros-positivos)
<i>mu2=0.85</i>	(média da distribuição beta para os verdadeiros-negativos)
<i>rho=0.70</i>	(correlação)
<i>shape1p=(mu1*(1-rho))/rho</i>	(parâmetro de forma para os verdadeiros-positivos)
<i>shape2p=(1-mu1)*(1-rho)/rho</i>	(parâmetro de forma para os verdadeiros-positivos)
<i>shape1n=mu2*(1-rho)/rho</i>	(parâmetro de forma para os verdadeiros-negativos)
<i>shape2n=(1-mu2)*(1-rho)/rho</i>	(parâmetro de forma para os verdadeiros-negativos)
<i>vp=rbetabinom.ab(n, size=m, shape1p, shape2p)</i>	(geração de verdadeiros-positivos)
<i>vn = rbetabinom.ab(n, size=m, shape1n, shape2n)</i>	(geração de verdadeiros-negativos)
<i>ni = rep(m,n)</i>	(m fixo)

Estimação da sensibilidade e especificidade:

sens=sum(vp)/sum(ni)
esp=sum(vn)/sum(ni)