UNIVERSIDADE FEDERAL DA PARAÍBA
CENTRO DE CIÊNCIAS EXATAS E DA NATUREZA
DEPARTAMENTO DE ESTATÍSTICA

Michelle Valeriano de Lima

Modelagem da Variável Citopatologia Anterior da Cidade de Recife, ano de 2013 via Regressão Logística Multinomial

João Pessoa, Fevereiro de 2015

MICHELLE VALERIANO DE LIMA

MODELAGEM DA VARIÁVEL CITOPATOLOGIA ANTERIOR DA CIDADE DE RECIFE, ANO 2013 VIA REGRESSÃO LOGÍSTICA MULTINOMIAL

Monografia apresentada ao Curso de Graduação em Estatística da Universidade Federal da Paraíba, como requisito parcial para obtenção do Grau de Bacharel. Área de Concentração: Estatística Aplicada.

Orientador(a): Profª. Drª. MARIA LÍDIA COCO TERRA.

João Pessoa, Fevereiro de 2015

MICHELLE VALERIANO DE LIMA

MODELAGEM DA VARIÁVEL CITOPATOLOGIA ANTERIOR DA CIDADE DE RECIFE, ANO 2013 VIA REGRESSÃO LOGÍSTICA MULTINOMIAL

Monografia apresentada ao Curso de Graduação em Estatística da Universidade Federal da Paraíba, como requisito parcial para obtenção do Grau de Bacharel. Área de Concentração: Estatística Aplicada.

Aprovada em Fevereiro de 2015.

BANCA EXAMINADORA

Prof.ª Dra. MARIA LÍDIA COCO TERRA - Orientador(a)
UFPB

Prof. Dr. JOÃO AGNALDO NASCIMENTO
UFPB

Prof. Dr. HEMÍLIO FERNANDES CAMPOS COÊLHO
UFPB

Dedico carinhosamente, este trabalho e todos os frutos
que eu obtiver, ao meu querido e amado pai (in memorian) que infelizmente
não conseguiu chegar a ver à realização do nosso sonho
te amarei por toda eternidade, com amor dedico.

Agradecimentos

A Deus, por sempre está ao meu lado, e nunca me deixando perder a fé.

Aos meus pais, em especial ao meu amado pai (in memorian), por tudo, conselhos, discussões, puxões de orelhas, e por sempre acreditar em mim.

As minhas amigas do coração, Paula e Lissandra, por todo apoio, conforto, amizade e companheirismo, nunca me abandonaram em momento algum em minha vida. Amo muito vocês meninas.

Ao meu amado namorado, Helivaldo, por todo amor, força, amizade, companheirismo e nos momentos de tristeza sempre me confortando.

As minhas amigas de escola Janielly, Vanessa, Priscilla, Julinez, Laís, Natália, Grayce e Amandy, por toda força, amizade, e amor a mim dedicado.

A professora Maria Lídia, pela excelente orientação, paciência, tranquilidade e sabedoria que teve, para lidar em todas as situação em que passamos nesses meses. Serei sempre grata.

Ao Professor Hemílio, pela oportunidade, contribuição acadêmica, incentivo e paciência. Sempre orientando majestosamente e sempre disponível, nas mais difícies situações.

As professoras, Andréia e Gilmara com tive os melhores momentos da minha vida, pela contribuição acadêmica e incentivo.

Ao professor João Agnaldo, a quem tenho profundo apreço e respeito, pois sempre acreditou em mim.

Aos Professores, pelas sugestões de melhorias deste trabalho.

Aos meus amigos, Saul, Jodavid, Alisson, que considero meus irmãos de coração, pela amizade, consideração, apoio e incentivo. Amizade que leverei por toda vida.

A Andreza, a quem sempre serei grata, pois me ensinou muito durante toda minha graduação, a quem amo como uma irmã.

A Marina, a quem dedico profundo respeito pela pessoa maravilhosa e profissional competente que você é, pelas parcerias, discussões, carinho, apoio e amizade.

A esses dois tesouros, que encontrei na Estatística, Maizza e Lígia, a quem dedico profundo carinho, respeito, consideração. Guardarei essa amizade para sempre.

A Aldine, amiga querida e companheira. Esteve comigo nos melhores e piores momentos da minha vida, nunca me abandonando.

A todos os Professores do DE-UPPB, por contribuírem na minha formação acadêmica.

Aos demais colegas da Graduação em Estatística, Marília, Geisislane, Ianne, Jéssica, Camila Ravena, Camila Ribeiro, Elaine, Henrique, Ramon e Pedro.

A todos os funcionários do Departamento de Estatística.

É melhor morrer amando
do que jamais ter amado.(Meu pai)

Resumo

O câncer de colo de útero, também chamado de cervical, é causado pela infecção persistente por alguns tipos (chamados oncogênicos) do Papilomavírus Humano - HPV, no entanto, fatores como início precoce da atividade sexual, multiplicidade de parceiros sexuais, uso de contraceptivos orais, tabagismo, situação conjugal e baixa condição socioeconômica têm sido apontados como fatores de risco importante para o desenvolvimento dessa neoplasia. O Papanicolau ou citopatológico é o modo de rastreamento para câncer de colo do útero mais utilizado na maior parte dos países do mundo. A Organização Mundial de Saúde (OMS) recomenda a realização do exame a cada três anos em mulheres com idade entre 25 a 65 anos após dois exames negativos com intervalo anual. O Sistema de Informação do Câncer do Colo do Útero - SISCOLO foi desenvolvido pelo INCA, em 1999, em parceria com o Departamento de Informática do SUS (Datasus), como ferramenta de gerência das ações do programa de controle do câncer de colo do útero o sistema está implantado nos laboratórios de citopatologia que realizam o exame citopatológico do colo do útero pelo Sistema Único de Saúde (módulo do prestador de serviço) e nas coordenações estaduais, regionais e municipais de detecção precoce do câncer (módulo de coordenação). Através do SISCOLO foi possível obter a variável de interesse para este trabalho é denominada de Citopatologia Anterior, em que, essa variável indica a frequência relacionada a realização de exames preventivos anteriores. Consequentemente, empregar uma regressão logística multinomial, devido a variável ser de natureza qualitativa, portanto adequada para o uso desde modelo da regressão logística multinomial. Foi possível constatar que as pacientes que "não sabe" se realizaram exames anteriormente são de maior frequência no intervalo de resultado entre 11 a 20 dias e para todos os níveis de escolaridade exceto ignorado \branco. Para as pacientes que "não informaram" se tinham realizado exames anteriormente são de maior frequência em todos os níveis da variável cor e raça, exceto para o nível "sem informação" e por fim, as pacientes que realizaram exames anteriormente possuem uma frequência equilibrada para os dois níveis da variável normalidade dos resultados. Com este trabalho, foi possível con-

cluir que o formulário de requisição do exame preventivo é um tanto quanto falho, pois não possuí variáveis importantes como *renda*, não sendo possível saber em quais condições socio-econômicas as pacientes vivem e o *número de vezes que a mesma paciente realizou o exame preventivo naquele ano*, com essa variável já seria possível detectar possíveis casos de câncer do colo do útero.

Palavras-chave: Exame Preventivo, Câncer de Colo de Útero, SISCOLO.

Abstract

Cervical cancer, also called cervical, is caused by persistent infection with certain types (called oncogenic) of human papillomavirus - HPV, however, factors such as early sexual activity, multiple sexual partners, use of oral contraceptives , smoking, marital status, and low socioeconomic status have been identified as important risk factors for the development of this neoplasm. The Pap test or Pap smear is the tracking mode for cervical cancer more used in most countries of the world. The World Health Organization (WHO) recommends the exam every three years for women aged 25 to 65 years after two negative tests with annual range. The Cervix Cancer Information System Uterus - SISCOLO was developed by INCA, at 1999, in partnership with the Department of SUS (Datasus) as management tool of control of program actions of cervical cancer the system is deployed in cytopathology laboratories performing Pap screening for cervical by the Unified Health System (service provider module) and the state coordination, regional and municipal early detection of cancer (coordination module). Through SISCOLO was possible to obtain the variable of interest for this work is called the Previous Cytopathology, wherein the frequency-related variable indicates the completion of previous screening tests. Consequently, employing a multinomial logistic regression, because the variable is qualitative, therefore suitable for use from the multinomial logistic regression model. It was found that patients who "do not know" took place earlier examinations are most often the result range between 11 to 20 days and for all educational levels except ignored \blank. For patients that "not reported" had taken place earlier examinations are more often at all levels of the variable color and race, except for the level "no information" and finally, patients who underwent tests have previously a balanced frequency for the two levels of the normal variable results. With this study, we conclude that the application form of the screening test is somewhat flawed because it does not possess important variables such as *income*, it is not possible to know which socioeconomic conditions patients live and the *number of times the same patient underwent the screening test that year*, with this variable would be possible to detect possible

cases of cervical cancer.

Keywords: Preventive examination, Colo Cancer Uterus, SISCOLO.

Conteúdo

Lista de Figuras

Lista de Tabelas

Capítulo 1

Introdução

O câncer do colo do útero, também chamado de cervical, é causado pela infecção persistente por alguns tipos (chamados oncogênicos) do Papilomavírus Humano - HPV. A infecção genital por este vírus é muito frequente, mas não causa doença na maioria das vezes. Entretanto, em alguns casos, podem ocorrer alterações celulares que poderão evoluir para o câncer. Essas alterações das células são descobertas facilmente no exame preventivo o Papanicolau (ou citopatológico). (INCA, 2014)

O exame preventivo Papanicolau ou citopatológico é o modo de rastreamento para câncer do colo do útero mais utilizado na maioria dos países do mundo. É um método sem dor e de baixo custo constituído basicamente de raspagem de tecido da ectocérvice e endocérvice do colo do útero por espátula de Ayres e escova citológica, respectivamente. O exame de Papanicolau e o tratamento do carcinoma *in situ* e de lesões de alto potencial de malignidade podem ser responsáveis pela redução de cerca de 80% da mortalidade por câncer de colo de útero. O Ministério da Saúde (MS) e a Organização Mundial de Saúde (OMS) preconizam que pelo menos 85% das mulheres realizem esse exame. (LEITE et al, 2010)

O Brasil foi um dos países precursores na utilização da citologia no diagnóstico do câncer do colo do útero. Há referência de que, em 1942, Antonio Vespasiano Ramos apresentou tese de docência intitulada "*Novo método de diagnóstico precoce do câncer uterino*", que se acredita ser o primeiro registro da utilização da citologia no diagnóstico do câncer no país. Além do pioneirismo, ao longo dos anos, o país vem ampliando a cobertura populacional aos exames citopatológicos. Estudos elaborados nos anos 80, considerando o número de exames citopatológicos realizados anualmente em relação ao número de mulheres que deveriam ser atingidas, estimaram coberturas nacionais de $1,2\%$ para o ano de 1984 e $7,7\%$ para 1985. Em 1994, um inquérito populacional realizado pelo Instituto Brasileiro de Opinião Pública e Estatística (Ibope) nas cinco macrorregiões do país mostrou coberturas que variaram entre 58% e 69%. (THULER, 2008)

O Brasil e suas regiões apresentaram índices de amostras insatisfatórias abaixo do limiar de 5% preconizado pela OMS. Entretanto, se considerado que o Brasil tem realizado, em média, dez milhões de exames citopatológicos por ano, o percentual de amostras insatisfatórias representa um quantitativo significativo de mulheres repetindo citologia. Acrescente-se a isso o fato de que 8% e 2% dos exames insatisfatórios, após a repetição, podem apresentar lesão intraepitelial ou câncer, respectivamente. (DIAS et al, 2010)

A Organização Mundial de Saúde (OMS) recomenda a realização do exame citológico a cada três anos em mulheres com idade entre 25 a 65 anos após dois exames negativos com intervalo anual.(CORREA et al, 2012). O exame preventivo quando realizado precocemente permite a cura de aproximadamente 100% dos casos. Por ano, o câncer do colo do útero, faz 4.800 vítimas fatais e 18.430 novos casos; o Instituto Nacional do Câncer estima que surjam 15.590 novos casos em 2014.

A infecção pelo papiloma vírus humano (HPV) tem sido apontada como principal fator de risco para o câncer de colo de útero, no entanto, fatores como início precoce da atividade sexual, multiplicidade de parceiros sexuais, uso de contraceptivos orais (pílulas do dia seguinte e anteconcepcional), tabagismo, situação conjugal e baixa condição sócio econômica têm sido apontados como fatores de risco importante para o desenvolvimento dessa neoplasia (ALBURQUERQUE et al, 2008).

Objetivos

O presente trabalho tem por objetivo modelar a variável citopatologia anterior através do modelo de Regressão Logística Multinomial.

Resumidamente,

- Traçar um perfil das mulheres que realizam o exame preventivo;
- Sugestões de melhorias no instrumento de coleta utilizado pelo SISCOLO, inserir no instrumento de coleta, variáveis como *renda*, sendo possível assim, traçar um perfil socioeconômico das pacientes que realizam o exame preventivo e o *número de vezes que a mesma paciente realizou o exame preventivo naquele ano*, com está variável seria possível detectar possíveis casos de câncer do colo do útero;
- Verificar quais níveis dos fatores influenciam na citopatologia anterior e para qual nível da mesma.

Capítulo 2

Desenvolvimento

2.1 Fonte dos Dados

O banco de dados foi obtido através do SISCOLO (Sistema de Informação do Câncer do Colo do Útero), em que, dispõe de informações relacionadas ao exame citopatológico cérvico-vaginal e microflora câncer do colo do útero realizados pelas mulheres da cidade do Recife com dados consolidados em 2013.

O SISCOLO foi desenvolvido pelo INCA em 1999, em parceria com o Departamento de Informática do SUS (Datasus), como ferramenta de gerência das ações do programa de controle do câncer de colo do útero. Os dados gerados pelo sistema permitem avaliar a cobertura da população-alvo, a qualidade dos exames, a prevalência das lesões precursoras, a situação do seguimento das mulheres com exames alterados, dentre outras informações relevantes ao acompanhamento e melhoria das ações de rastreamento, diagnóstico e tratamento. O sistema está implantado nos laboratórios de citopatologia que realizam o exame citopatológico do colo do útero pelo Sistema Único de Saúde (módulo do prestador de serviço) e nas coordenações estaduais, regionais e municipais de detecção precoce do câncer (módulo de coordenação). (SISCOLO, 2014)

2.2 Descrição dos Dados

O banco de dados é formado por 96401 pacientes do sexo feminino que realizaram o exame preventivo citopatológico contra o câncer do colo do útero na cidade de Recife, capital do Estado de Pernambuco no período de janeiro a dezembro de 2013.

Este banco de dados é formado por dez variáveis, em que, foram consideradas apenas sete variáveis, incluindo a variável resposta. As variáveis que formam o banco de dados, são

descritas a seguir juntamente com suas respectivas abreviaduras, que serão utilizadas no decorrer da apresentação do trabalho:

Intervalo de Coleta (“intecolet”) - período em dias que a paciente realiza o exame preventivo. Varia entre 10 a 30 dias;

Intervalo de Resultado (“intresult”) - período em dias que a paciente receberá o resultado do exame preventivo realizado. Varia entre 10 a 30 dias;

Tempo de Exame (“tempexam”) - indica o tempo que o laboratório leva para liberar os resultados do exame preventivo realizado. Varia entre 30 a 60 dias;

Faixa Etária (“faixetar”) - indica a frequência por faixa etária das pacientes que realizaram o exame preventivo. Varia entre 11 a 60 anos de idade;

Cor e Raça (“corraca”) - indica a etnia da paciente que realizou o exame preventivo. Classificados em: branca, preta, amarela, parda, indígena e sem informação;

Citopatologia Anterior (“citopatant”) - frequência com que a paciente realizou o exame preventivo. Divido em: não, não sabe, sim e não informado;

Normalidade dos Resultados (“normal”) - indica se os resultados obtidos foram normais ou não. Caso o resultado tenha acusado normalidade dos resultados, a paciente repetirá o exame preventivo após um ano. No entanto, se o resultado não for normal, então a paciente deverá repetir o exame daqui a seis meses;

Escolaridade (“esc”) - nível de instrução da paciente. Classificado em: ignorado \branco, analfabeta, ensino fundamental incompleto, ensino fundamental completo, ensino médio incompleto, ensino médio completo, ensino superior completo;

Adequabilidade da Amostra (“adeq”) - indica se a amostra coletada foi adequada ou não. É considerada insatisfatória quando material acelular ou hipocelular (ou seja, menos de 10% do esfregaço) ou leitura prejudicada (mais de 75% do esfregaço) por presença de: sangue, piócitos, artefatos de dessecamento, contaminantes externos ou intensa superposição celular. Deve-se repetir o exame preventivo entre seis a doze semanas após o resultado de amostra não adequada.(INCA, 2014)

Tempo Último Preventivo (“temp”) - esta variável indica a periodicidade em anos que a paciente realizou seu último exame preventivo. Classificado em ignorado \branco, a cada um,

dois, três, quatro e cinco anos.

Vale ressaltar que as variáveis ***faixa etária, tempo último preventivo e adequabilidade da amostra*** não entraram na modelagem, devido a quantidade de observações contidas em seus vetores que diferem das demais variáveis.

2.3 Metodologia

Segundo Figueira (2006), considera-se uma coleção de $r+1$ variáveis independentes denotadas por $\underset{\sim}{X} = (X_0, X_1, ..., X_r)$, onde $\overline{x} = (x_0, x_1, \ldots, x_r)$ com $x_0 = 1$ e uma variável aleatória(v.a.) Y de natureza nominal que pode assumir os níveis $\{0, 1, \ldots, q\}$. Descreve-se o *logit*, comparando com o $Y = k$ com $Y = 0$ para $k \in \{1, \ldots, q\}$. O valor *zero* então é denominado de *categoria de referência*.

As funções g_k são *preditores lineares* que se ligam as variáveis explicativas atráves da função *logit*. Essas funções g_k são definidas como:

$$\begin{aligned} g_k \equiv g_k(x) &= \ln\left[\frac{P(Y=k|x)}{P(Y=0|x)}\right] \\ g_k &= \beta_{k0}x_0 + \beta_{k1}x_1 + \cdots + \beta_{kr}x_r \\ g_k &= x'\beta_k, \quad k \in \{0, \ldots, q\} \end{aligned} \tag{2.1}$$

onde

$$\beta_k = (\beta_{k0} + \cdots + \beta_{kr})' x_{k0} = 1 \tag{2.2}$$

Assumindo n observações independentes de Y, denotadas por $y_1, \ldots, y_n$, associadas aos valores de $x_i = (x_{i0}, \ldots, x_{ir})$ para $i \in 1, \ldots, n$, o *logit*, dado pela equação (2.1) é apresentado como:

$$\begin{aligned} g_{k1} \equiv g_{k1}(x_1) &= \beta_{k0}x_{10} + \beta_{k1}x_{11} + \cdots + \beta_{kr}x_{x1r} + \varepsilon_1 \\ g_{k2} \equiv g_{k2}(x_2) &= \beta_{k0}x_{20} + \beta_{k2}x_{21} + \cdots + \beta_{kr}x_{x2r} + \varepsilon_2 \\ &\vdots \\ g_{kn} \approx g_{kn}(x_n) &= \beta_{k0}x_{n0} + \beta_{k1}x_{n1} + \cdots + \beta_{kr}x_{nr} + \varepsilon_n, \end{aligned} \tag{2.3}$$

onde $x_{xi0} = 1$, para $i \in \{1, \ldots, n\}$ e os erros, ε_i que seguem as seguintes suposições para todo

$i, l \in \{1, \ldots, n\}$

$$\begin{aligned} E(\varepsilon_i|x_i) &= 0, \\ Var(\varepsilon_i|x_i) &= Var(Y_i|x_i), \\ Cov(\varepsilon_i, \varepsilon_l) &= 0, \quad \text{se } i \neq l. \end{aligned} \tag{2.4}$$

As v.a.'s $Y_1, \ldots, Y_n$ satisfazem um modelo multinomial se uma amostra de tamanho um de cada Y_i pode ser expressa como:

$$\pi_{ki} \equiv \pi_{ki}(x_i) = \frac{exp(g_{ki})}{1 + exp(g_{ki})}, \tag{2.5}$$

onde g_{ki} é obtido pela expressão (2.1), para qual x_{ij} é uma constante conhecida e β_{kj} é um parâmetro desconhecido, os erros ε_i e $\pi_{ki}(x)$ representa $P = (Y_i = k|x)$, com $i \in \{1, \ldots, n\}$, $j \in \{0, \ldots, r\}$ e $k \in \{0, \ldots, q\}$.

Da expressão (2.1), em que, $\exp[g_{0i}(x)] = 1$, e desta forma $\beta_{0j} = 0$, para qualquer $j \in \{0, \ldots, r\}$, e para cada nível da v.a.'s Y pode assumir $r+1$ coeficientes, ou seja, o modelo apresenta um total de $q(r+1)$ coeficientes.

A probabilidade condicional para um modelo com $q+1$ categorias é dada por

$$P(Y = k|x) = \frac{\exp[g_k(x)]}{\sum_{k=0}^{q} \exp[g_k(x)]}, \tag{2.6}$$

onde $g_k(x)$ é dada pela equação (2.1), para $k \in \{1, \ldots, q\}$ e $g_0(x) = 0$.

Demonstração: Pelo uso das propriedades de logaritmo na expressão (2.1), obtemos:

$$\exp[g_k(x)] = \frac{P(Y = k|x)}{P(Y = 0|x)}, \tag{2.7}$$

para $k \in \{0, \ldots, q\}$, ou ainda

$$P(Y = k|x) = \exp[g_k] P(Y = 0|x)$$

Pela propriedade da probabilidade total, tem-se que:

$$\sum_{k=1}^{q} P(Y = k|x) = 1. \tag{2.8}$$

Agora substituindo (2.7) na (2.8), tem-se:

$$\sum_{k=1}^{q} \exp[g_k(x)] P(Y = 0|x) = 1,$$

em que, por propriedade de somatório, obtém-se:

$$P(Y=0|x) = \frac{1}{\sum_{k=1}^{q} \exp[g_k(x)]}, \tag{2.9}$$

Ao substituir o resultado obtido na expressão (2.9) em (2.7), obtêm-se a seguinte afirmação. Como T é uma v.a. de natureza politômica, com $q+1$ valores, expressa-se uma observação como $y = \pi(x) + \varepsilon$, e assim, a v.a. ε pode assumir $q+1$ valores. Se $y_i = k$ então $\varepsilon_i = k = \pi(x_i)$ com probabilidade $P = (Y = k|x_i)$, para qualquer $k \in \{0,\ldots,q\}$ e $i \in \{1,\ldots,n\}$.

A v.a. ε tem distribuição de probabilidade multinomial com média zero e variância igual a v.a. Y.

Demonstração: A esperança de ε dado x_i,

$$\begin{aligned}
E(\varepsilon|x_i) &= \sum_{k=0}^{q} \varepsilon_k P(\varepsilon = \varepsilon_k|x_i) \\
&= -\pi(x_i)P(Y=0|x_i) + \cdots + (q-\pi(x_i))P(Y=q|x_i) \\
&= -\pi(x_i)\sum_{k=0}^{q} P(Y=k|x_i) + \sum_{k=0}^{q} kP(Y=k|x_i) \\
&= -\pi(x_i) + E(Y|x_i) = -\pi(x_i) + \pi(x_i) = 0.
\end{aligned}$$

Agora, obtendo a variância condicional de ε,

$$\begin{aligned}
Var(\varepsilon|x_i) &= \sum_{k=0}^{q} \varepsilon_k^2 P(\varepsilon = \varepsilon_k|x_i) = \sum_{k=0}^{q} (k-\pi(x_i))^2 P(Y=k|x_i) \\
&= \sum_{k=0}^{q} (k^2 - 2k\pi(x_i) + \pi(x_i)^2)P(Y=k|x_i) \\
&= \sum_{k=0}^{q} k^2 P(Y=k|x_i) - 2\pi(x_i)\sum_{k=0}^{q} kP(Y=k|x_i) + \pi(x_i)^2 \sum_{k=0}^{q} P(Y=k|x_i) \\
&= E(Y^2|x_i) - 2\pi(x_i)E(Y|x_i) + \pi(x_i)^2 \\
&= E(Y^2|x_i) - \pi(x_i) = Var(\varepsilon|x_i).
\end{aligned}$$

2.4 Função de Verosssimilhança

Para se construir a função de verossimilhança é necessário introduzir $q+1$ variáveis auxiliares com o objetivo de simplificar a notação utilizada, mas que não se emprega em nenhuma análise posterior. As variáveis auxiliares são apresentadas da seguinte forma.

$$\begin{aligned}
&Y = 0, \text{ então } Y_0 = 1, Y_1 = 0, \ldots, Y_q = 0 \\
&Y = 1, \text{ então } Y_0 = 0, Y_1 = 1, \ldots, Y_q = 0
\end{aligned}$$

De forma geral,

$$Y = k, \text{então} \quad Y_k = 1, Y_l = 0, \quad \text{para } i \neq l \in 0, 1, \ldots, q.$$

Com isso, pode-se definir uma matriz auxiliar Q como

$$Q = \begin{bmatrix} 1 & 0 & \cdots & 0 \\ 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 1 \end{bmatrix}_{(q+1)\times(q+1)} \tag{2.10}$$

onde, os elementos q_{rs} da linha r correspondem, respectivamente, aos valores assumidos pelas variáveis auxiliares Y_k, com $r \in \{1, \ldots, q\}$ e $k = r - 1$, quando $Y = k$. Não importa qual seja o valor assumido por Y, o somatório $\sum_{k=1}^{q} Y = k$ sempre será igual a um.

Nos modelos de regressão logística, um dos métodos utilizado para estimação dos parâmetros é o da máxima verossimilhança, pois os métodos de mínimos quadrados quando aplicado a um modelo, em que, a variável resposta é nominal, os estimadores obtidos não possuem as mesmas propriedades.

A função de verossimilhança $\ell(\beta)$ para uma amostra de n observações independentes é dada por

$$\ell(\beta) = \prod_{i=1}^{n} [\pi_0(x_i)^{Y_{0i}} \pi_1(x_i)^{Y_{1i}} \ldots \pi_q(x_i)^{Y_{qi}}], \tag{2.11}$$

onde $\pi = \pi(x_i)$, $x_i = (x_{i0}, \ldots, x_{ir})$ e $i \in \{1, \ldots\}$.

Demonstração: É comum se utilizar da função log de verossimilhança, obtida após aplicação do logaritmo natural em ambos os lados da expressão (2.10), assumindo a forma

$$L(\beta) = \ln \left\{ \prod_{i=1}^{n} [\pi_0(x_i)^{Y_{0i}} \pi_1(x_i)^{Y_{1i}} \ldots \pi_q(x_i)^{Y_{qi}}] \right\}. \tag{2.12}$$

Seja β o vetor de parâmetros relacionados com a probabilidade $P(Y_i = k|x_i)$, para $i \in \{1, \cdots, n\}$ e $k \in \{0, \ldots, q\}$. Então o estimador de β, pelo método de máxima verossimilhança denotado por $\hat{\beta}$, é a solução das equações

$$\frac{\partial L(\beta)}{\partial \beta_{kj}} = \sum_{i=1}^{n} x_{ij}(y_{ki} - \pi_{ki}), \tag{2.13}$$

para $k \in \{1, \cdots, q\}$, $j \in \{0, \cdots, r\}$ e $\pi_{ki} = \pi_k(x_i)$, com $x_{0i} = 1$, para qualquer i.

Demonstração: Ao aplicar as propriedades de somatório e logaritmo na função apresentada em (2.12), tem-se

$$L(\beta) = \sum_{i=1}^{n} \left\{ \sum_{k=1}^{q} y_{ki} g_k(x_i) - \ln \left[\sum_{k=1}^{q} \exp[g_k(x_i)] \right] \right\}. \tag{2.14}$$

As equações de verossimilhança (2.13) são obtidas através das primeiras derivadas parciais de (2.14) com respeito a cada um dos $q(r+1)$ parâmetros desconhecidos. Para simplificar a notação, define-se $\pi_{ki} = \pi_k(x_i)$. Assim sendo, a forma geral das equações apresentadas em (2.14) é

$$\frac{\partial L(\beta)}{\partial \beta_{kj}} = \sum_{i=1}^{n} x_{ij}(y_{ki} - \pi_{ki}),$$

para $k \in \{1, \cdots, q\}$, $j \in \{1, \cdots, r\}$ e $x_{0i} = 1$ para qualquer $i \in \{1, \cdots, n\}$. O estimador de máxima verossimilhança, $\hat{\beta}$, é obtido igualando ambos os lados a zero e resolvendo o sistema para β.

Para obter a significância dos $q(r+1)$ coeficientes no modelo apresentado na expressão (2.5), o teste da razão da verossimilhança é baseado na estatística G (menos duas vezes a verossimilhança sem a variável no modelo divido pela verossimilhança com a variável no modelo) e apresenta distribuição assintótica qui-quadrado com $q(r+1) - r$ graus de liberdade.

Assumindo o contexto da expressão (2.5), o teste da razão de verossimilhança de tamanho α é dado por

$$\begin{aligned} &H_0 : \beta = B \\ &H_1 : \beta \neq B, \quad \text{para } B \in M_{(q+1)\times(r+1))} \end{aligned} \tag{2.15}$$

em que rejeita-se a hipótese nula, se $P\left(\chi^2_{q(r+1)-r} > G\right) < \alpha$, e $M_{(q+1)\times(r+1))}$ representa o conjunto de todas as matrizes de dimensão $(q+1) \times (r+1)$.

A matriz das segundas derivadas parciais é necessária para se obter a matriz de informação de *Fisher*, $I(\hat{\beta})$, e o estimador da matriz de variância de variâncias-covariâncias para $\hat{\beta}$. A expressão geral, dos elementos na matriz das segundas derivadas parciais é

$$\frac{\partial^2 L(\beta)}{\partial \beta_{kj} \partial \beta_{kl}} = \sum_{i=1}^{n} x_{il} x_{ij} \pi_{ki}(1 - \pi_{ki}), \tag{2.16}$$

para $k, m \in \{1, \ldots, q\}$ e $j, l \in \{0, 1, \ldots, r\}$. A matriz $I(\hat{\beta})$, de ordem $2(r+1)$, possui elementos que são simétricos aos valores encontrados nas expressões (2.14)quando avaliados em $\hat{\beta}$.

O estimador da matriz de variâncias-covariâncias de $\hat{\beta}$ é a inversa da matriz da informação, ou seja,

$$\widehat{Var}(\hat{\beta}) = I(\hat{\beta})^{-1}. \tag{2.17}$$

2.5 Teste de Wald

Com base no resultado obtido na expressão (2.17), pode-se apresentar o análogo multinomial ao Teste de Wald. A estatística desde teste é dada pela expressão

$$W = \hat{\beta}'[I(\hat{\beta})]\hat{\beta}'.$$

Sabe-se que sob $H_0 : \beta = 0$, a estatística W possui distribuição qui-quadrado com $q(r+1) - r$ graus de liberdade. E, da mesma forma que seus análogos, este teste não apresenta vantagens computacionais sobre o teste da razão de verossimilhança.

2.6 Intervalos de Confiança

Os cálculos dos intervalos de confiança utilizados para os coeficientes β_{kj}, com $k \in \{0, \cdots, q\}$ e $j \in \{1, \cdots, r\}$, de forma análoga ao caso binário.

O intervalo a $100(1-\alpha)\%$ de confiança para β_{kj} é dado por

$$\left[\hat{\beta}_{kj} - z_{\frac{\alpha}{2}}\widehat{SE}(\hat{\beta_{kj}}), \hat{\beta_{kj}} + z_{\frac{\alpha}{2}}\widehat{SE}(\hat{\beta_{kj}})\right],$$

onde $z_{\frac{\alpha}{2}}$ é o quantil de uma normal padrão dado por $P(z > z_{\frac{\alpha}{2}}) = \frac{\alpha}{2}$ e $\widehat{SE}(\hat{\beta_{kj}})$ representa o estimador do desvio padrão de $\hat{\beta}_{kj}$. O estimador de $\widehat{SE}(\hat{\beta}_{kj})$ é a raíz quadrada do elemento da k-ésima linha e j-ésima coluna da matriz $I(\hat{\beta})^{-1}$.

Para a obtenção dos intervalos de confiança para os estimadores das funções *logits*, $g_k(x)$, com $k \in \{0, \ldots, q\}$. Expressão o *logit* em sua notação dados por (2.3).

O estimador da variância de $\hat{g}(x)$, representado por $\widehat{Var}[\hat{g}_k(x)]$, requer a obtenção da variância da soma, resultando em

$$\widehat{Var}[\hat{g}_k(x)] = x'\widehat{Var}(\hat{\beta}_k)x \tag{2.18}$$

O estimador $\widehat{Var}[\hat{g}_k(x)]$ pode ser obtido através do *método Delta*, que fornece valores dos desvios-padrão de estatísticas que podem ser representadas como função de outras estatísticas que possuem distribuição assintótica conjunta normal. A formalização do método *Delta* pode ser obtida em Agresti (1984) e Agresti (1990).

O intervalo a $100(1-\alpha)\%$ de confiança para $g_k(x)$, com $k \in \{0, \cdots, q\}$ é dado por

$$\left[\hat{g}_k(x) - z_{\frac{\alpha}{2}}\sqrt{\widehat{Var}[\hat{g}_k(x)]}, \hat{g}_k(x) + z_{\frac{\alpha}{2}}\sqrt{\widehat{Var}[\hat{g}_k(x)]}\right],$$

onde $\widehat{Var}[\hat{g}_k(x)]$ é dada pela expressão (2.18) e $z_{\frac{\alpha}{2}}$ é o quantil de uma normal padrão dado por $P(z > z_{\frac{\alpha}{2}}) = \frac{\alpha}{2}$.

O intervalo a $100(1-\alpha)\%$ de confiança para $\pi_k(x)$ é dado por

$$\left\{ \frac{\exp\left[\hat{g}_k(x) - z_{\frac{\alpha}{2}}\sqrt{\widehat{Var}[\hat{g}_k(x)]}\right]}{1+\exp\left[\hat{g}_k(x) - z_{\frac{\alpha}{2}}\sqrt{Var}[\hat{g}_k(x)]\right]}, \frac{\exp\left[\hat{g}_k(x) + z_{\frac{\alpha}{2}}\sqrt{\widehat{Var}[\hat{g}_k(x)]}\right]}{1+\exp\left[\hat{g}_k(x) + z_{\frac{\alpha}{2}}\sqrt{Var}[\hat{g}_k(x)]\right]} \right\},$$

onde $\widehat{Var}[\hat{g}_k(x)]$ é dada pela expressão (2.18) e $z_{\frac{\alpha}{2}}$ é o quantil de uma normal padrão dado por $P(z > z_{\frac{\alpha}{2}}) = \frac{\alpha}{2}$.

Capítulo 3

Resultados e Discussões

A Figura 3.1 abaixo apresenta o histograma da variável resposta citopatologia anterior, frequência com que a paciente realizou o exame preventivo, que é dividida em quatro níveis: “não”, “não sabe”, “sim” e “não informado”. É possível observar que o terceiro nível (”sim”) é o que possuí maior frequência, ou seja, há um registro maior de pacientes que realizaram o exame citopatológico anteriormente.

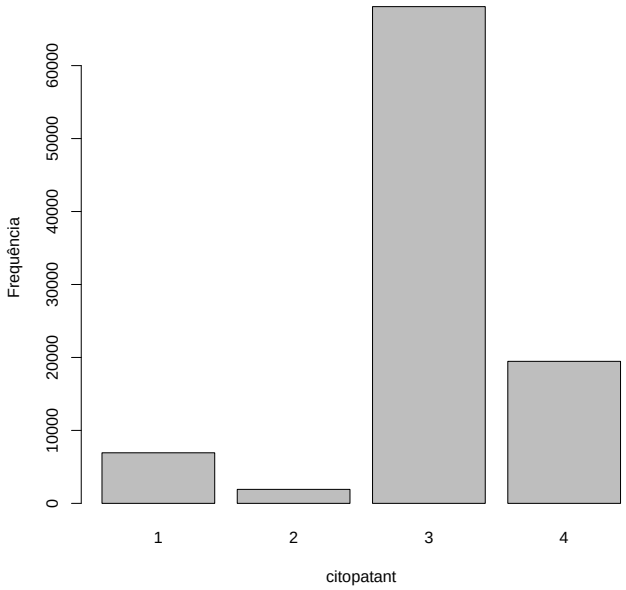


Figura 3.1: Gráfico de Barras da Variável Resposta Citopatologia Anterior.

O conjunto dos gráficos de barras das variáveis explicativas do modelo escolhido apresentados na Figura 3.2, mostram que para a variável intervalo de resultado, na qual possuí quatro níveis, em que variam entre 10 a 30 dias, o nível com maior frequência é o segundo, o que indica que os resultados do exame com maior frequência variam entre 11 a 20 dias. O gráfico de barras da variável escolaridade que possuí seis níveis (ignorado\branco, analfabeta, ensino

fundamental incompleto, ensino fundamental completo, ensino médio completo e ensino superior completo), em que, o nível com maior frequência é o primeiro, ou seja, as pacientes que realizaram o exame citopatológico possuem o nível de escolaridade ignorado\branco. O gráfico de barras da variável cor e raça possuí seis níveis (branca, preta, parda, amarela, indígena, sem informação), o nível com maior frequência é o sexto, cuja, as pacientes que realizaram o exame citopatológico não se tem informações sobre a cor e a raça das mesmas. O gráfico de barras da variável normalidade, possuí dois níveis (sim, não), cujo nível de maior frequência é o segundo, ou seja, a maioria das pacientes que realizaram o exame citopatológico, não tiveram seus exames dentro da normalidade, o que implica a paciente realizar nova o exame após seis meses da realização do último exame.

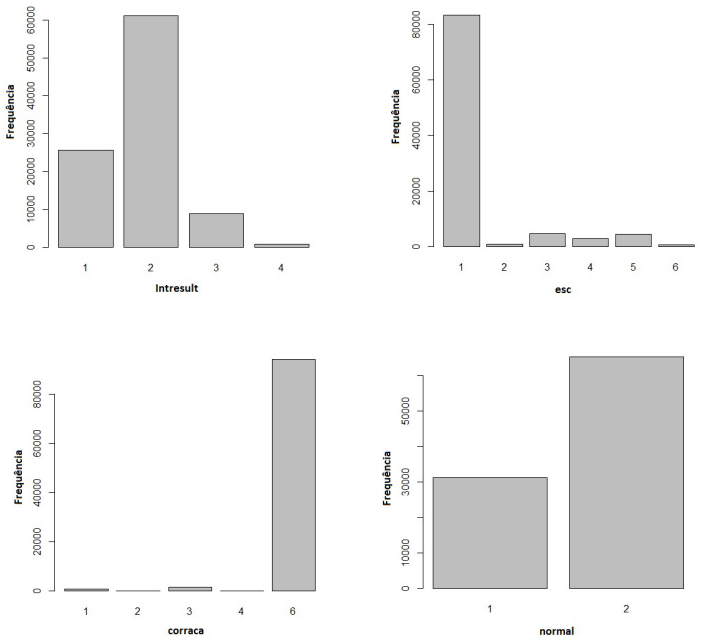


Figura 3.2: Gráficos de Barras das Variáveis Explicativas.

A escolha do modelo mais adequado se deu pelo método de seleção *stepwise* com as variáveis explicativas intervalo de coleta (**intecolet**), intervalo de resultado (**intresult**), tempo de exame (**tempexam**), escolaridade (**esc**), cor e raça (**corraca**) e normalidade dos resultados (**normal**) e como critério de seleção do modelo mais adequado utilizou-se o menor valor do AIC (Critério de Informação de Akaike).

O modelo escolhido possuí as variáveis intervalo de resultado (**intresult**), escolaridade (**esc**), cor e raça (**corraca**) e normalidade dos resultados (**normal**). O teste Qui-Quadrado foi realizado com o intuito de verificar se o modelo selecionado está bem ajustado, para isso se compara os resultados do desvio em relação ao quantil da distribuição qui-quadrado, em que, o desvio deverá ser menor que o quantil para se aceitar que o modelo proposto está bem ajustado. A Tabela 3.1 apresenta os resultados obtidos para a verificação do ajuste para o modelo proposto.

Como o desvio foi menor que o quantil, logo o modelo proposto está bem ajustado.

Tabela 3.1: Verificação da bondade do ajuste do modelo.

Desvio	quantil
0.6801828	0.9949088

A Tabela 3.2 apresenta os resultados do modelo proposto. Os coeficientes dos parâmetros e os erros padrão, que são estimativas para os desvios padrão das distribuições dos coeficientes e seus valores permitem medir a confiança estatística que se pode ter com relação a essas estimativas, ou seja, quanto menor o erro padrão, maior a confiança que se pode ter nos modelos ajustados. Os erros padrão são obtidos através da raíz quadrada dos termos da diagonal da matriz de covariância dos coeficientes.

Tabela 3.2: Resultados do modelo escolhido pelo método de seleção *stepwise* para o ajuste dos dados.

Variável	Coeficiente	Erro Padrão
Intercepto	-39.49	0.32
intresult	-1.73	6.53
esc	0.80	3.97
corraca	-0.15	2.63
normal	27.86	2.61
AIC	73538.84	
Deviance	73508.84	

Para $Y = 4$, ou seja, o nível "não informado" da variável resposta, o intervalo de confiança de 95% para os parâmetros do modelo escolhido encontram-se na Tabela 3.3. Por exemplo, o intervalo de confiança para o parâmetro da variável **corraca** é $[-5.29; 4.99]$. Para os demais níveis da variável resposta, foram realizados intervalos de confiança, porém os valores obtidos dos intervalos de confiança não foram próximos ao valor dos coeficientes, portanto não foi colocado neste trabalho.

Tabela 3.3: Intervalos de confiança dos parâmetros do modelo escolhido.

Parâmetros	IC [2.5%; 97.5%]
Intercept	-40.11; -38.86
intresult	-14.52; 11.06
esc	-6.97; 8.58
corraca	-5.29; 4.99
normal	22.73; 32.97

A Figura 3.3 exibe a proporção de níveis entre a variável resposta citopatologia anterior com as variáveis explicativas intervalo de resultado e escolaridade, respectivamente. Para a variável explicativa intervalo de resultado, a proporção das pacientes que "não sabem" se realizaram exames preventivos anteriormente cresce a partir do primeiro nível e após do segundo nível mantem-se constante. A proporção das pacientes que "não sabe" se realizaram exames preventivos anteriormente, cresce a partir do primeiro nível de escolaridade e após o segundo nível mantem-se proporcionalmente constante para os demais níveis de escolaridade.

A Figura 3.4 evidência a proporção de níveis entre a variável resposta com as variáveis explicativas cor e raça e normalidade dos resultados, nessa ordem. A proporção de pacientes que "não informaram" se realizaram exames anteriormente é proporcionalmente constante nos primeiros quatro níveis da variável explicativa cor e raça e após o quarto nível decaí para o quinto nível. A proporção de pacientes que tenham realizado exame anteriormente manteve-se constante para ambos os níveis da variável explicativa normalidade dos resultados.

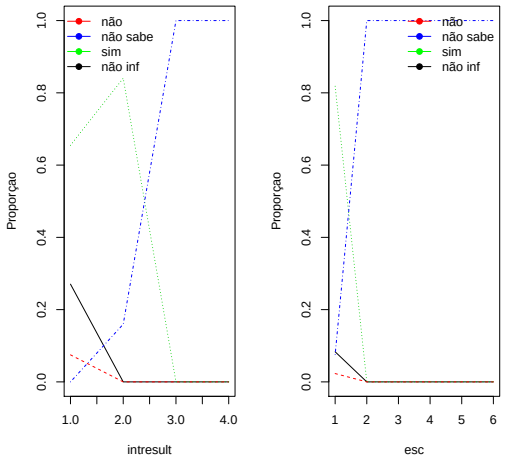


Figura 3.3: Gráficos da proporção da variável resposta com as variáveis explicativas Intervalo de Resultado e Escolaridade.

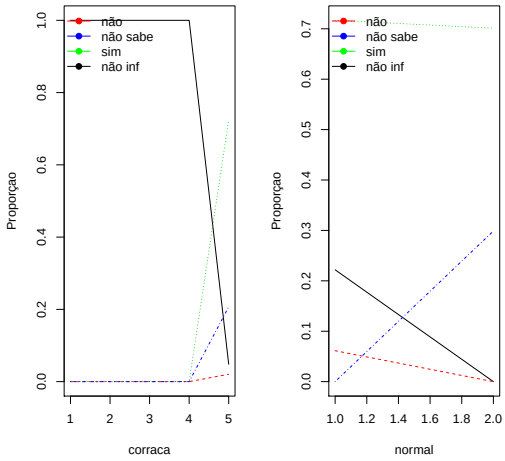


Figura 3.4: Gráficos da proporção da variável resposta com as variáveis explicativas Cor e Raça e Normalidade dos Resultados.

Capítulo 4

Conclusões

Com a distribuição multinomial foi possível modelar a variável resposta citopatologia anterior, variável correlacionada ao exame preventivo cérvico-vaginal e microflora câncer do colo do útero, realizado no ano de 2013 por pacientes do sexo feminino na cidade do Recife. De todas as variáveis explicativas do banco de dados, as variáveis que ajudaram a explicar a variável resposta citopatologia anterior (não, não sabe, sim e não informação) foram: intervalo de resultado (0 a 10, 11 a 20, 21 a 30, maior que 30 dias), escolaridade (ignorado\branco, analfabeta, ensino fundamental incompleto, ensino fundamental completo, ensino médio incompleto, ensino médio completo, ensino superior completo), cor e raça (branca, preta, parda, amarela, indígena, sem informação) e normalidade dos resultados (sim e não). Todas possuem níveis, com isso foi possível observar quais níveis para cada variável explicativa são mais significativa para cada nível da variável resposta.

O intervalo de resultado do nível entre 11 a 20 dias foi mais significativo com o nível não sabe da variável resposta. Dessa forma, as pacientes que realizaram o exame preventivo anterior, nesse intervalo de resultado não sabem se realizaram o exame anteriormente.

A escolaridade para os níveis analfabeta, ensino fundamental incompleto, ensino fundamental completo, ensino médio completo e ensino superior foram todos significativos para o nível não sabe da variável resposta, ou seja, para quase todos os níveis de escolaridade as pacientes que realizaram o exame preventivo não têm conhecimento de exames anteriormente realizados.

Cor e raça nos níveis branca, preta, parda, amarela e indígena foram mais significativas com o nível não informados da variável resposta, isto é, quase todos os níveis da cor e raça não possuem informação de exames preventivos realizados anteriormente.

A normalidade dos resultados foi significativa para seus dois únicos níveis sim e não com o nível sim da variável resposta, assim as pacientes que realizaram o exame preventivo anterior já haviam realizado o exame anteriormente e ocorreu um equilíbrio entre as pacientes que tiveram resultados normais e não normais.

O instrumento utilizado para a coleta destas informações é ainda um tanto falho, pois

deixa de lado algumas variáveis que seriam interessantes como a *renda*, com ela seria possível traçar um perfil socioeconômico da paciente, possibilitando assim, saber quais as condições financeira das pacientes. Outra variável também importante seria o *número de vezes a mesma paciente realizou o exame preventivo*, em que, para isso seria necessário a geração de um banco de dados para cada município que realizam o exame preventivo, com a finalidade de montar um cadastro de cada paciente que realizou o exame e quantas vezes a mesma paciente o fez e o ano do exame realizado, sendo assim possível detectar possíveis casos de câncer do colo do útero.

A variável cor e raça indígena, não há informação de mulheres que tenham feito o exame preventivo, pois não há comunidades indígena próxima a cidade do Recife. Também destaca-se a cor e raça negra, essas pacientes realizam poucos exames, sendo da mesma forma necessário políticas públicas para alcançar esse público específico de mulheres.

4.1 Sugestões para Trabalhos Futuros

Para a realização de trabalhos futuros alguns sugestões serão feitas a seguir:

- Realizar outra análise com outras variáveis do banco de dados;
- Dicotomizar a variável resposta e fazer uma regressão logística, no caso, utilizaria os níveis com maior frequência da variável reposta;
- Acrescentar no formulário de requisição do exame citopatológico, as variáveis *renda*, com isso seria possível traçar um perfil socioeconômico das pacientes que realizam o exame preventivo e a variável *número de vezes que a paciente realizou o exame preventivo naquele ano*, sendo possível detectar possíveis casos de câncer do colo do útero.

Referências

[1] AGRESTI, A. Analysis of Ordinal Categorical Data. New York: John Wiley.

[2] AGRESTI, A. Analysis of Ordinal Categorical Data. New York: John Wiley.

[3] ALBUQUERQUE, K.M.; FRIAS, P.G.; ANDRADE, C.L.T.; AQUINO, E.M.L.; MENEZES, G.; SZWARCWALD, C.L. Cobertura do teste de Papanicolau e fatores associados à não-realização: um olhar sobre o Programa de Prevenção do Câncer do Colo do Útero em Pernambuco, Brasil. Cad. de Saúde Pública, Rio de Janeiro, RJ, vol.25, sup.2, pag. S301-S309, dez. de 2008.

[4] CORREA, M.S.; SILVEIRA, D.S.; SIQUEIRA, F.V.; FACCHINI, L.A.; PICCINI, R.X.; THUMÉ, E.; TOMASI, E. Cobertura e adequação do exame citopatológico de colo uterino em Estados da Regiões Sul e Nordeste do Brasil. Cad. Saúde Pública, Rio de Janeiro, RJ, nº28, vol.12, p.2257-2266, dez. de 2012.

[5] DIAS, M.B.K.; TOMAZELLI, J.G.; ASSIS, M. Rastreamento do câncer de colo do útero no Brasil: análise de dados do Siscolo no período de 2002 a 2006. Disponível em : <http://scielo.iec.pa.gov.br> Acesso em 13 de janeiro de 2015.

[6] FARAWAY, J.J. Extending the linear model with R: Generalized Linear, Mixed Effects and Nonparametric Regression Models. Chapman & Hall/ CRC, Boca Raton, Flórida, 2006.

[7] FIGUEIRA, C.V. Modelos de Regressão Logística. Universidade Federal do Rio Grande do Sul, Porto Alegre - Instituto de Matemática, 31 de Março.

[8] INCA, Disponível em: <http://www2.inca.gov.br> Acesso em 13 de novembro de 2014.

[9] LEITE, F.M.C.; AMORIM, M.H.C.; NASCIMENTO, L.G.D.; MENDONÇA, M.R.F.; GUEDES, N.S.A.; TRISTÃO, K.M. Mulheres submetidas à coleta de Papanicolau: perfil socioeconômico e reprodutivo. Revista Brasileira de Pesquisa em Saúde, Espírito Santo, ES, ano 1, v.12, p. 57-62, mar. de 2010.

[10] SISCOLO, Disponível em: < http:// w3.datasus.gov.br> Acesso em 10 de outubro de 2014.

[11] THULER, L.C.S. Mortalidade por câncer do colo do útero no Brasil. Revista Brasileira de Ginecologia e Obstetrícia, Rio de Janeiro, RJ, n^{o}30, p.216-218, mar. de 2008.

Apêndice A

Comandos do R utilizados nas análises do presente trabalho

```
## PACOTES
require(foreign)
require(nnet)
require(ggplot2)
require(reshape2)

## VETORES
# intervalo de coleta
intecolet=c(rep(1,90843), rep(2,4232), rep(3,645), rep(4,681))
intecolet=as.vector(intecolet)

# Intervalo do resultado
intresult=c(rep(1,25622), rep(2,61057), rep(3,8894), rep(4,828))
intresult=as.vector(intresult)

# Tempo do exame
tempexam=c(rep(1,92234), rep(2,3645), rep(3,522))
tempexam=as.vector(tempexam)

# Escolaridade
esc=c(rep(1,83202), rep(2,771), rep(3,4664),
rep(4,2866), rep(5,4346), rep(6,552))
esc=as.vector(esc)

# Cor e Raca
corraca=c(rep(1,728), rep(2,87), rep(3,1582), rep(4,15), rep(6,93989))
corraca=as.vector(corraca)

# Faixa Etaria
faixetar=c(rep(1,30), rep(2,468), rep(3,6149), rep(4,9061), rep(5,10775),
 rep(6,11640), rep(7,11063), rep(8,11233), rep(9,10424), rep(11,6471),
```

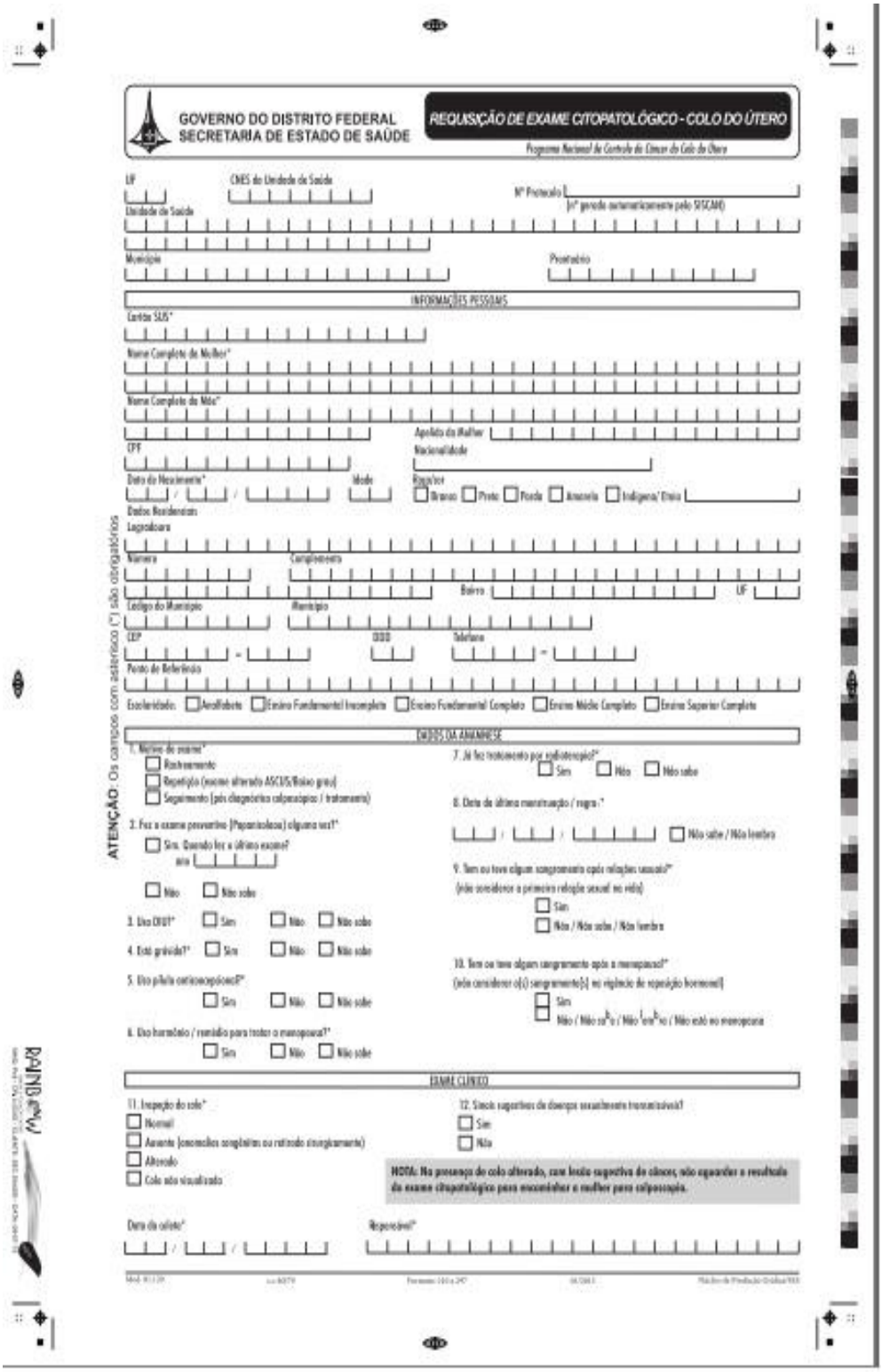
```
 rep(12,4704), rep(13,5444))
faixetar=as.vector(faixetar)

# Citopatologia anterior
citopatant=c(rep(1,6932), rep(2,1917), rep(3,68085), rep(4,19467))
citopatant=as.vector(citopatant)

# Normalidade dos resultados
normal=c(rep(1,31245), rep(2,65156))
normal=as.vector(normal)

banco <- cbind(citopatant,intecolet,intresult,tempexam,esc,corraca,normal)
banconew=data.frame(banco)
attach(banconew)

## MODELO GERAL
model<-multinom(citopatant ~ intresult + esc + corraca + normal+
tempexam+intresult+intecolet, data=banconew)

## AJUSTE DO MODELO MAIS ADEQUADO
model1<-step(model)
model1<-multinom(citopatant ~ intresult + esc + corraca + normal,
data=banconew)

summary(model1)

## Estatistica de teste - Wald
z <- summary(model1)$coefficients/summary(model1)$standard.errors
z

## ESTATISTICA DEVIANCE
deviance(model1)-deviance(model)
pchisq(0.6801828, model$edf-model1$edf, lower=F)


## PLOTS DA VARIAVEL RESPOSTA COM CADA VARIAVEL EXPLICATIVA

pdf("plots.pdf")
par(mfrow = c(1,2))
matplot(prop.table(table(banconew$intresult,banconew$citopatant),1),
 type="l", xlab="intresult", ylab="Proporçao",
 color=c("red","blue","green","black") )
legend(x="topleft", c("não","não sabe", "sim", "não inf"),border="white",
col = c("red","blue","green", "black"), lty = 1, bty="n", pch=19)

matplot(prop.table(table(banconew$esc,banconew$citopatant),1), type="l",
 xlab="esc",  ylab="Proporçao", color=c("red","green","blue","black"))
legend(x="topright", c("não","não sabe", "sim", "não inf"),
```

```
border="white", col = c("red","blue","green", "black"), lty = 1,
bty="n", pch=19)

dev.off()

pdf("plots2.pdf")
par(mfrow = c(1,2))
matplot(prop.table(table(banconew$corraca,banconew$citopatant),1),
type="l", xlab="corraca", ylab="Proporçao",
color=c("red","green","blue","black"))
legend(x="topleft", c("não","não sabe", "sim", "não inf"),
border="white", col = c("red","blue","green", "black"), lty = 1,
 bty="n", pch=19)

matplot(prop.table(table(banconew$normal,banconew$citopatant),1), type="l",
 xlab="normal", ylab="Proporçao", color=c("red","green","blue","black"))
legend(x="topleft", c("não","não sabe", "sim", "não inf"),border="white",
col = c("red","blue","green", "black"), lty = 1, bty="n", pch=19)

dev.off()


## ANALISE DISCRITIVA DO MODELO AJUSTADO
hist(citopatant, xlab="citopatant", ylab="Frequência", main="")

par(mfrow = c(2,2))
hist(intresult, xlab="intresult", ylab="Frequência", main="")
hist(esc, xlab="esc", ylab="Frequência", main="")
hist(corraca,xlab="corraca", ylaba="Frequência", main="")
hist(normal, xlab="normal", ylaba="Frequência", main="")


## INTERVALO DE CONFIANÇA
confint(model1)
```

Apêndice B

Formulário de Requisição de Exame Citopatológico - Colo do Útero

APÊNDICE B. FORMULÁRIO DE REQUISIÇÃO DE EXAME CITOPATOLÓGICO - COLO DO ÚTERO



GOVERNO DO DISTRITO FEDERAL
SECRETARIA DE ESTADO DE SAÚDE

REQUISIÇÃO DE EXAME CITOPATOLÓGICO - COLO DO ÚTERO

Programa Nacional de Controle do Câncer do Colo do Útero

UF | CNES da Unidade de Saúde | Nº Protocolo (nº gerado automaticamente pelo SISCAN)

Unidade de Saúde

Município | Prontuário

INFORMAÇÕES PESSOAIS

Cartão SUS*

Nome Completo da Mulher*

Nome Completo da Mãe*

CPF | Apelido da Mulher | Nacionalidade

Data do Nascimento* | Idade | Raça/cor: ☐ Branca ☐ Preta ☐ Parda ☐ Amarela ☐ Indígena/ Etnia

Dados Residenciais
Logradouro

Número | Complemento | Bairro | UF

Código do Município | Município

CEP | DDD | Telefone

Ponto de Referência

Escolaridade: ☐ Analfabeta ☐ Ensino Fundamental Incompleto ☐ Ensino Fundamental Completo ☐ Ensino Médio Completo ☐ Ensino Superior Completo

DADOS DA ANAMNESE

ATENÇÃO: Os campos com asterisco (*) são obrigatórios

1. Motivo do exame*
☐ Rastreamento
☐ Repetição (exame alterado ASC/US/Baixo grau)
☐ Seguimento (pós diagnóstico colposcópico / tratamento)

2. Fez o exame preventivo (Papanicolaou) alguma vez?*
☐ Sim. Quando foi a última exame? ano
☐ Não ☐ Não sabe

3. Usa DIU?* ☐ Sim ☐ Não ☐ Não sabe

4. Está grávida?* ☐ Sim ☐ Não ☐ Não sabe

5. Usa pílula anticoncepcional?* ☐ Sim ☐ Não ☐ Não sabe

6. Usa hormônio / remédio para tratar a menopausa?* ☐ Sim ☐ Não ☐ Não sabe

7. Já fez tratamento por radioterapia?* ☐ Sim ☐ Não ☐ Não sabe

8. Data da última menstruação / regra:* / / ☐ Não sabe / Não lembra

9. Tem ou teve algum sangramento após relações sexuais?* (não considerar o primeiro relação sexual na vida)
☐ Sim
☐ Não / Não sabe / Não lembra

10. Tem ou teve algum sangramento após a menopausa?* (não considerar o(s) sangramento(s) na vigência de reposição hormonal)
☐ Sim
☐ Não / Não sabe / Não lembra / Não está na menopausa

EXAME CLÍNICO

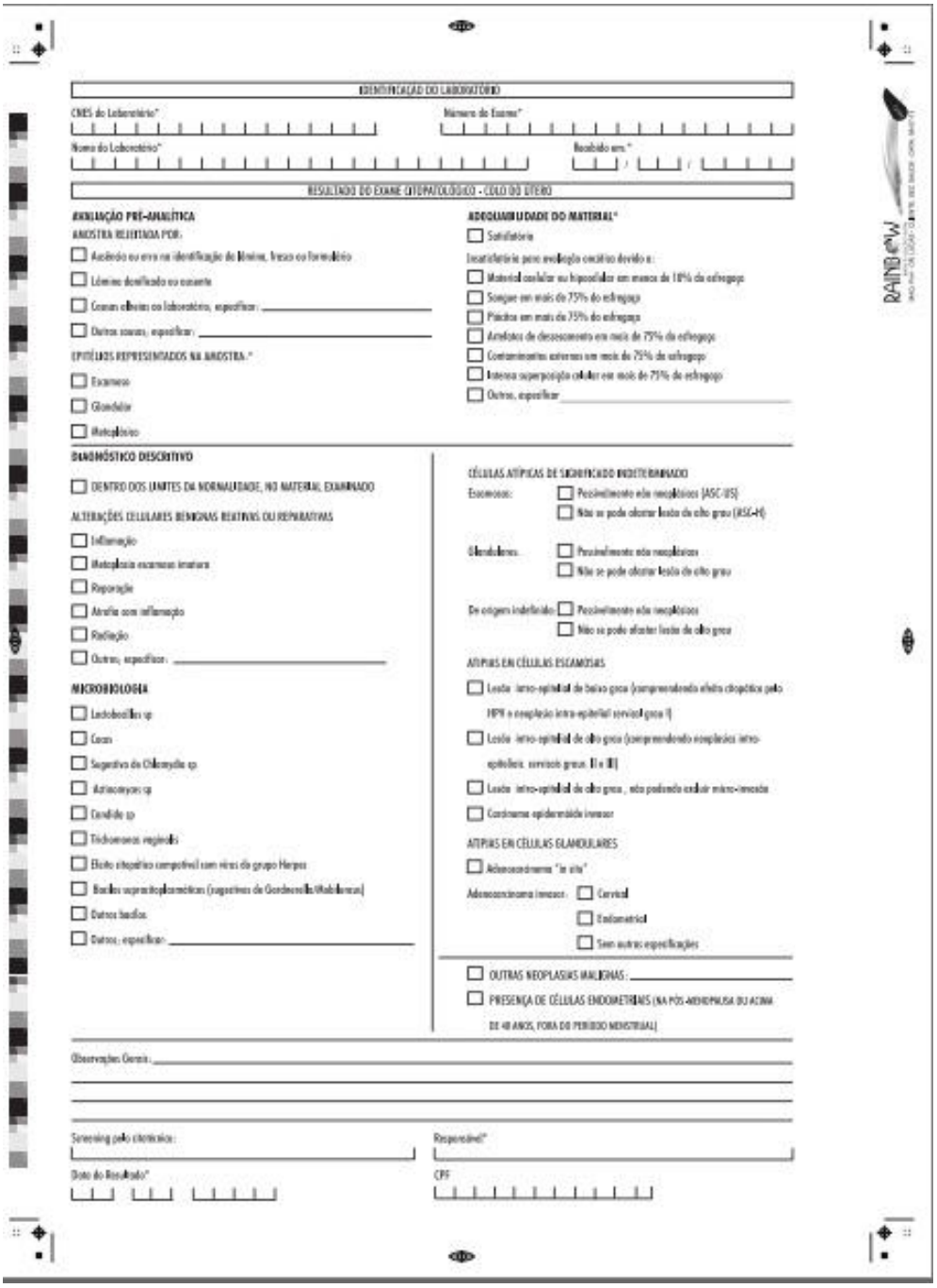
11. Inspeção do colo*
☐ Normal
☐ Ausente (cervicectomia congênitas ou retirada cirurgicamente)
☐ Alterado
☐ Colo não visualizado

12. Sinais sugestivos de doenças sexualmente transmissíveis?
☐ Sim
☐ Não

NOTA: Na presença de colo alterado, com lesão sugestiva de câncer, não aguardar o resultado do exame citopatológico para encaminhar a mulher para colposcopia.

Data da coleta* / / | Responsável*

APÊNDICE B. FORMULÁRIO DE REQUISIÇÃO DE EXAME CITOPATOLÓGICO - COLO DO ÚTERO



IDENTIFICAÇÃO DO LABORATÓRIO

CNES do Laboratório*

Número do Exame*

Nome do Laboratório*

Recebido em:* / /

RESULTADO DO EXAME CITOPATOLÓGICO - COLO DO ÚTERO

AVALIAÇÃO PRÉ-ANALÍTICA

AMOSTRA REJEITADA POR:

- ☐ Ausência ou erro na identificação da lâmina, frasco ou formulário
- ☐ Lâmina danificada ou ausente
- ☐ Causas alheias ao laboratório, especificar: _____
- ☐ Outras causas, especificar: _____

EPITÉLIOS REPRESENTADOS NA AMOSTRA:*

- ☐ Escamoso
- ☐ Glandular
- ☐ Metaplásico

ADEQUABILIDADE DO MATERIAL*

- ☐ Satisfatório

Insatisfatório para avaliação citológica devido a:

- ☐ Material celular ou hipocelular em menos de 10% da esfregaço
- ☐ Sangue em mais de 75% do esfregaço
- ☐ Piócitos em mais de 75% do esfregaço
- ☐ Artefatos de dessecamento em mais de 75% do esfregaço
- ☐ Contaminantes externos em mais de 75% do esfregaço
- ☐ Intensa superposição celular em mais de 75% do esfregaço
- ☐ Outros, especificar _____

DIAGNÓSTICO DESCRITIVO

- ☐ DENTRO DOS LIMITES DA NORMALIDADE, NO MATERIAL EXAMINADO

ALTERAÇÕES CELULARES BENIGNAS REATIVAS OU REPARATIVAS

- ☐ Inflamação
- ☐ Metaplasia escamosa imatura
- ☐ Reparação
- ☐ Atrofia com inflamação
- ☐ Radiação
- ☐ Outras, especificar: _____

MICROBIOLOGIA

- ☐ Lactobacilos sp
- ☐ Cocos
- ☐ Sugestivo de Chlamydia sp
- ☐ Actinomyces sp
- ☐ Candida sp
- ☐ Trichomonas vaginalis
- ☐ Efeito citopático compatível com vírus do grupo Herpes
- ☐ Bacilos supracitoplasmáticos (sugestivos de Gardnerella Vaginalis)
- ☐ Outros bacilos
- ☐ Outros, especificar: _____

CÉLULAS ATÍPICAS DE SIGNIFICADO INDETERMINADO

Escamosas:
- ☐ Possivelmente não neoplásicas (ASC-US)
- ☐ Não se pode excluir lesão de alto grau (ASC-H)

Glandulares:
- ☐ Possivelmente não neoplásicas
- ☐ Não se pode excluir lesão de alto grau

De origem indefinida:
- ☐ Possivelmente não neoplásicas
- ☐ Não se pode excluir lesão de alto grau

ATIPIAS EM CÉLULAS ESCAMOSAS

- ☐ Lesão intra-epitelial de baixo grau (compreendendo efeito citopático pelo HPV e neoplasia intra-epitelial cervical grau I)
- ☐ Lesão intra-epitelial de alto grau (compreendendo neoplasias intra-epiteliais cervicais graus II e III)
- ☐ Lesão intra-epitelial de alto grau, não podendo excluir micro-invasão
- ☐ Carcinoma epidermóide invasor

ATIPIAS EM CÉLULAS GLANDULARES

- ☐ Adenocarcinoma "in situ"

Adenocarcinoma invasor:
- ☐ Cervical
- ☐ Endometrial
- ☐ Sem outras especificações

- ☐ OUTRAS NEOPLASIAS MALIGNAS: _____
- ☐ PRESENÇA DE CÉLULAS ENDOMETRIAIS (NA PÓS-MENOPAUSA OU ACIMA DE 40 ANOS, FORA DO PERÍODO MENSTRUAL)

Observações Gerais: _____

Screening pelo citotécnico: _____

Responsável* _____

Data do Resultado* / /

CPF