

Keilane Alves Pereira

Associação entre a não realização do exame
preventivo de câncer do colo de útero e
características socioeconômicas, demográficas,
comportamentais e de saúde da mulher, usando
regressão logística binária.

Niterói – RJ, Brasil

2014



Universidade Federal Fluminense

Keilane Alves Pereira

Associação entre a não realização do exame preventivo de
câncer do colo de útero e características socioeconômicas,
demográficas, comportamentais e de saúde da mulher, usando
regressão logística binária.

Trabalho de Conclusão de Curso

Monografia apresentada para obtenção do grau de Bacharel em
Estatística pela Universidade Federal Fluminense.

Orientadora: Profa. Luz Amanda Melgar Santander

Co-orientador: Prof. José Rodrigo de Moraes

Niterói - RJ, Brasil

2014



Universidade Federal Fluminense

Keilane Alves Pereira

Associação entre a não realização do exame preventivo de câncer do colo de útero e características socioeconômicas, demográficas, comportamentais e de saúde da mulher usando regressão logística binária.

Monografia de Projeto Final de Graduação sob o título "*Associação entre a não realização do exame preventivo de câncer do colo de útero e características socioeconômicas, demográficas, comportamentais e de saúde da mulher usando regressão logística binária*", defendida por Keilane Alves Pereira e aprovada em 26 de junho de 2014, na cidade de Niterói, no Estado do Rio de Janeiro, pela banca examinadora constituída pelos professores:

Profa. Dra. Luz Amanda Melgar Santander
Orientadora
Departamento de Estatística – UFF

Prof. Dr. José Rodrigo de Moraes
Coorientador
Departamento de Estatística – UFF

Dra. Marcella de Oliveira Santos
Instituto Nacional do Câncer José Alencar Gomes da Silva – INCA

Prof. Dr. Carlos Augusto Faria
Departamento Materno-Infantil da Faculdade de Medicina - UFF

Niterói, 01 de julho de 2014.

Resumo

O câncer do colo do útero é um dos cânceres mais incidentes na população brasileira, podendo ser curável se diagnosticado no início. Assim a realização do exame preventivo para câncer do colo do útero (Papanicolaou) é uma das formas mais eficazes para o controle da doença, além de ser de baixo custo. Este trabalho buscou estabelecer associação entre a não realização do Papanicolaou e características socioeconômicas, demográficas, comportamentais e de saúde da mulher, usando o modelo de regressão logística binária. Os dados utilizados são da PNAD 2008, que é uma pesquisa amostral complexa realizada pelo IBGE. Entre os resultados de estudo, verificou-se uma maior chance de não realização do exame Papanicolaou para mulheres residentes na região Centro-Oeste (em relação ao Norte), na área rural, menos escolarizadas, com menores rendimentos domiciliares, que não praticam atividade física, para mulheres que nunca fumaram (em relação as ex-fumantes) e para mulheres sem nenhuma doença crônica (em relação a 3 ou mais doenças crônicas). Além disso, observou-se que a ausência de plano de saúde, a não realização de consulta médica, e do exame de mamografia, e o fato de não ter filho estão significativamente associadas com a chance da não realização do exame Papanicolaou. Estes resultados mostram a necessidade de sensibilizar determinados grupos de mulheres sobre a importância da realização de exames como o Papanicolaou, por meio de educação em saúde, e de universalizar o acesso aos serviços de saúde. Além disso, observa-se a importância de investir na capacitação profissional de atenção primária, e agilizar o diagnóstico tanto na rede pública quanto na rede privada.

Palavras-chaves: Saúde da Mulher, câncer de colo do útero, exame Papanicolaou, modelo logístico binário, Pesquisa Nacional por Amostras de Domicílios (PNAD) 2008.

*Dedico este trabalho primeiramente a Deus por
permitir que concluísse essa etapa da minha vida, e
também a Ana Maria minha mãe, Ademar meu pai, e
minha irmã Keila... A minha família.*

Agradecimentos

Agradeço em especial a minha orientadora Luz Amanda, não só por me orientar para a conclusão deste trabalho, mas por se dispor com sua amizade e compreensão nos momentos de dificuldade nessa reta final de UFF. Também agradeço ao meu coorientador José Rodrigo, pela paciência, dedicação e amizade que muito me ajudou para a realização deste projeto, e na criação do tema proposto para esta discussão. Agradeço muito aos dois por formarem uma excelente dupla de orientadores e se mostrarem disponíveis para marcar reuniões para este trabalho.

Agradeço aos professores que me conduziram nessa trajetória como futura estatística, em especial ao professor Licínio, um grande mestre que pude contar com sua amizade nos momentos de dificuldade na UFF; um grande professor dedicado aos alunos, e a arte de ensinar.

Aos funcionários Adelmo Bandeira, Alcir, Alexandre Martins, Beatriz Barbosa, Seu Cláudio, Wellington Batista que tiveram muita paciência, e me deram uma força, seja na hora de pedir uma ajuda com um arquivo que não abria ou ajudando a imprimir folhas...

Também agradeço a minha amiga de mais de 10 anos Suelen Rosales que me informou sobre a oportunidade de estagiar no INCA, instituição onde aprendi muito e fiz grandes amizades, como Rejane Reis, Julio Fernando, Juliana Moreira e a grande Marceli Santos que muito me influenciou a gostar da área de saúde. Muito obrigada a essa equipe que é dedicada ao importante trabalho que faz, e devo muito a eles por me ajudar no meu crescimento profissional, e também por darem o pontapé inicial para querer fazer este trabalho.

Com grande carinho agradeço aos meus amigos de estrada da UFF, que sempre estiveram ao meu lado estudando comigo, me acalmando nas provas, escutando as minhas lamentações, simplesmente sendo amigos... As pessoas que muito me incentivaram nessa minha longa trajetória na UFF, em especial, amigos e colegas como Ana Paula Cassani, Ulisses Silva, Paulo Chagas, Heloisa Ferreira, Vanessa Chaffin, Viviana Lobo, Camila Silva, Danielle Freitas, Bruno Lucian, Fabio Loureiro, Carolina Vital, Juliana Freitas, Evelynne Guimarães, Gabriel de Aguiar, Lucas Mattos, Luciana Moura, Luciana Leite, Nadine Melloni, Sergio Felipe, Clarissa Cardoso e Kiese

Quiavauca. Em especial agradeço a Fernanda Lima, minha amiga de todas as horas, com quem eu sempre pude contar.

Agradeço aos meus colegas de trabalho que me ajudaram quando precisei, Vanessa Felipe, Karina Marques, Lorena Tinoco, Ingo Gameleira, Iuri Miranda e Leonardo Menezes.

Com muita emoção agradeço aos meus familiares e amigos da vida, que me ajudaram economicamente e afetivamente a estudar na UFF, Marli Nesti, Bruno Pompeo, Luiz Antônio, Dona Dalva, Evelyn Ribeiro, Keila Pereira, Ademar Alves, Ana Maria.

Finalmente agradeço a quem me fez dar boas risadas e pode me deixar tranqüila em meio a confusão... a Alana.

Sumário

Lista de Figuras.....	2
Lista de Quadros.....	3
Lista de Tabelas.....	4
1. Introdução	5
1.1.Câncer na atualidade.....	5
1.2.Prevenção do câncer.....	5
1.3.Câncer do colo do útero e o exame Papanicolaou.....	7
2. Objetivos	10
2.1. Objetivo geral.....	10
2.2. Objetivos específicos.....	10
3. Material e Métodos	11
3.1. Pesquisa Nacional por Amostra de Domicílios (PNAD).....	11
3.1.1.Plano amostral da PNAD	11
3.2.População de estudo.....	13
3.3. Variáveis utilizadas no estudo.....	13
3.4. Modelo de regressão logística binária.....	15
3.4.1. Especificação do modelo.....	15
3.4.2. Métodos de estimação dos parâmetros do modelo.....	17
3.4.2.1. Método de máxima verossimilhança (MV).....	17
3.4.2.2. Método de máxima pseudo-verossimilhança (MPV).....	19
3.4.3. Inferência no modelo.....	22
3.4.3.1 Teste de Wald de significância geral dos parâmetros do modelo.....	22
3.4.3.2 Teste de Wald de significância individual dos parâmetros do modelo.....	23
3.4.4. Medidas de qualidade do ajuste (pseudo- R^2).....	24
3.4.5. Avaliação da capacidade preditiva do modelo.....	25
3.4.5.1.Taxa de classificações correta.....	25
3.4.5.2. Área sob a curva ROC.....	26
4. Resultados e Discussões.....	28
4.1. Resultados.....	28
4.2. Discussões.....	41

5. Conclusões e Considerações Finais.....	43
Referências Bibliográficas.....	44

Lista de Figuras

Figura 1: Exame Papanicolaou.....	8
Figura 2: Área sob a Curva ROC.....	40

Lista de Quadros

Quadro 1: Variáveis demográficas, socioeconômicas, comportamentais e de saúde relativas a mulheres de 25 anos ou mais de idade.....	14
Quadro 2: Distribuição dos indivíduos segundo as categorias observadas e previstas pelo modelo de regressão logística binária.....	25
Quadro 3: Classificação do poder discriminatório do modelo logístico binário baseada na área sob a curva ROC.....	27

Lista de Tabelas

Tabela 1: Distribuição percentual de mulheres com 25 anos ou mais de idade, segundo características demográficas e socioeconômicas no Brasil, e pelo percentual de condição de realização do exame preventivo para câncer de colo do útero (Papanicolaou).....29

Tabela 2: Distribuição percentual de mulheres com 25 anos ou mais de idade, segundo características comportamentais, de saúde e outras no Brasil, e pelo percentual de condição de realização do exame preventivo para câncer de colo do útero (Papanicolaou).....31

Tabela 3: Resultados dos testes de significância geral para os parâmetros do modelo logístico binário explicativo da não realização do exame Papanicolaou, considerando todas as variáveis explicativas (dezenove).....33

Tabela 4: Resultados do ajuste do modelo de regressão logística binária explicativo da não realização do exame Papanicolaou considerando todas as variáveis explicativas.....34

Tabela 5: Resultados dos testes de significância geral para os parâmetros do modelo logístico binário explicativo da não realização do exame Papanicolaou, considerando doze variáveis explicativas selecionadas.....36

Tabela 6: Resultados do ajuste do modelo de regressão logística binária explicativo da não realização do exame Papanicolaou considerando todas as doze variáveis explicativas selecionadas (modelo final)37

1. Introdução

1.1. Câncer na atualidade

O câncer é uma doença responsável por mais de 12% de todas as causas de óbito no mundo, onde mais de sete milhões de pessoas morrem por ano devido a essa doença (BRASIL, 2007). Este quadro deve-se à grande exposição dos indivíduos a fatores de risco cancerígenos e às mudanças demográficas, como a redução da mortalidade e natalidade, que contribuem para o aumento da expectativa de vida e o envelhecimento da população, que por sua vez, provocam o aumento da prevalência de doenças crônicas-degenerativas, entre elas o câncer (BRASIL, 2007).

No Brasil, vem ocorrendo transições no perfil epidemiológico, ou seja, mudanças nas causas de mortalidade e morbidade, em conjunto com outras transformações demográficas, sociais e econômicas.

No Brasil, estima-se que para o biênio 2014/2015, haverá 576 mil novos casos de câncer, sendo 190 mil no grupo das mulheres (sem considerar os casos de câncer de pele não melanoma). Neste grupo, os mais incidentes seriam os cânceres de pele não melanoma (83.710 mil casos novos), mama (57.120 mil), colo do útero (15.590 mil), cólon e reto (17.530 mil), pulmão (10.930 mil) e tireóide (8.050 mil) (INCA, 2014).

O câncer é responsável por grande parte da mudança do perfil de adoecimento da população brasileira. Atualmente, registra-se o aumento da incidência de cânceres de cólon e reto, mama e próstata, os quais estão associados ao melhor nível socioeconômico. Por outro lado, observam-se taxas de incidência elevadas de cânceres de colo do útero, estômago, cabeça e pescoço, que estão associados a condições sociais menos favorecidas (INCA, 2011).

1.2. Prevenção do câncer

A Organização Mundial de Saúde (OMS) considera que cerca de 40% das mortes por câncer poderiam ser evitadas, sendo a prevenção essencial para o controle do câncer. A prevenção do câncer consiste em um conjunto de medidas para reduzir ou

evitar a exposição a fatores que aumentam a possibilidade de um indivíduo desenvolver uma determinada neoplasia ou sofrer um determinado agravo, comumente chamado de fatores de risco.

Os fatores de risco de câncer podem ser encontrados no ambiente físico¹, ser herdados, ou representar comportamentos e costumes próprios de um determinado ambiente social e cultural².

O risco de câncer em determinada população depende das suas condições sociais, ambientais, políticas e econômicas, bem como das características biológicas e comportamentais dos indivíduos. Assim, o aparecimento de vários tipos de câncer deve-se não somente a susceptibilidade genética, mas também a sua interação com outros fatores ou condições relacionadas ao modo de vida e do ambiente (INCA, 2012). Desse modo, o Ministério da Saúde elaborou a Política Nacional de Prevenção e Controle do Câncer, cujo objetivo é reduzir a incidência e a mortalidade por câncer no Brasil, por meio de ações contínuas, que levem a conscientização da população quanto aos fatores de risco para o câncer, que promovam a detecção precoce dos tumores passíveis de rastreamento e propiciem o acesso a um tratamento equitativo e de qualidade em todo território nacional (INCA, 2003b).

De acordo com as recomendações governamentais em vigor (portarias ministeriais), a implementação de políticas de controle do câncer implica organização de linhas de cuidado que passem todos os níveis de atenção e modalidades de atendimento, em um modelo assistencial que articule recursos, garantindo acesso aos serviços e tratamento necessários.

Em 2006, o Pacto pela Saúde, aprovado pelo Conselho Nacional de Saúde e publicado na Portaria GM/MS nº399, de 22 de fevereiro de 2006, considerou em uma de suas dimensões (“Pacto pela Vida”), o controle do câncer do colo do útero e de mama entre as prioridades de saúde no Brasil (INCA, 2012).

¹ Ambiente em geral: ocupacional (indústrias químicas e afins), de consumo (alimentos, medicamentos), etc.

² Estilo e hábitos de vida, tais como tipo de alimentação, prática de exercícios físico, consumo de bebidas alcoólicas e etc.

1.3. Câncer do colo do útero e o exame Papanicolaou

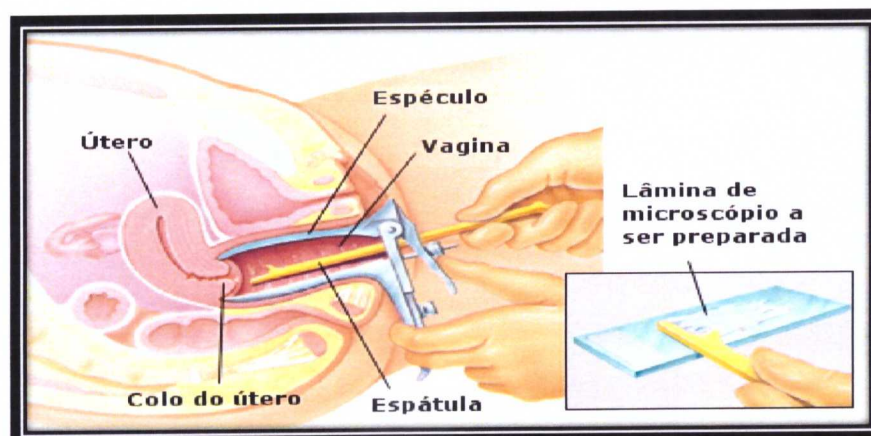
O câncer do colo do útero corresponde a cerca de 15% de todos os tipos de câncer em mulheres, sendo atualmente o segundo câncer mais comum no mundo. Cerca de 80% dos tumores de colo do útero ocorrem nos países em desenvolvimento, ocupando a primeira posição em alguns deles.

O câncer do colo do útero inicia-se a partir de uma lesão pré-invasiva, curável em até 100% dos casos (anormalidades epiteliais conhecidas como Displasias ou Neoplasias Intraepiteliais Cervicais – NIC), que geralmente progride lentamente, por anos, antes de atingir o estágio invasor da doença, tornando a cura mais difícil, quando não impossível (WHO, 1988).

Estudos indicam que a presença de Papiloma Vírus Humano (HPV)³ contribui para o aparecimento do câncer do colo do útero, mas não constitui uma causa suficiente (THULER *et al.*, 2012). Uma das medidas de prevenção primária do câncer do colo do útero pode ser realizada através do uso de preservativos durante a relação sexual (não é uma medida 100% segura para evitar o contágio pelo HPV). Outras medidas de prevenção primária se referem as seguintes práticas: manter cuidados higiênicos; ter parceiro sexual fixo (ou ao menos um número reduzido); visitar regularmente o ginecologista, assim como o parceiro sexual deve procurar o médico, para realizar exames de prevenção. Não fumar e não realizar uso de contraceptivos orais hormonais por período prolongado (5 anos ou mais, em mulheres com HPV) são outras medidas para evitar o desenvolvimento do câncer do colo do útero (PRADO *et al.*, 2012; THULER *et al.*, 2012; ALDRIGHI *et al.*, 2002).

A forma mais eficaz para o controle do câncer do colo de útero continua sendo o rastreamento por meio do exame preventivo de Papanicolaou, que consiste na coleta de material para exame na parte externa (ectocérvix) e interna (endocérvix) do colo do útero (Figura 1). O material coletado é afixado em lâmina de vidro, corado pelo método de Papanicolaou e então examinado ao microscópio. É um exame rápido, de custo relativamente baixo, e efetivo para a detecção precoce de lesões pré-cancerosas do colo

³ HPV – Papiloma Vírus Humano, vem do inglês Human Papiloma Virus. É um vírus que infecta a pele e possui mais de 200 variações diferentes, e a maioria aparece através de verrugas, que pode ter sintomas ou não, infecções subclínicas (não visíveis).



do útero.

Figura 1: Exame Papanicolaou.

Como o câncer do colo do útero tem uma longa fase pré-clínica, se o exame de Papanicolaou for repetido periodicamente, espera-se que as lesões precursoras que não tenham sido identificadas num primeiro exame sejam diagnosticadas em exames subsequentes. A evolução do câncer do colo do útero ocorre, em geral, de modo assintomático; somente em alguns casos há sintomas, como sangramento após a realização de relação sexual; dor pélvica e secreção vaginal com odor fétido (BRASIL, 2003).

A estratégia utilizada nas últimas décadas, em diversos países, para a detecção precoce do câncer de colo de útero e suas lesões precursoras é a realização do exame citológico de Papanicolaou. Este exame consiste em uma tecnologia simples, eficaz e de baixo custo para o sistema de saúde, podendo ser realizado por um profissional de saúde treinado adequadamente, sem a necessidade de uma infra-estrutura sofisticada (OLIVEIRA et al., 2006). Ele possibilita um rastreamento de até 80% dos casos de câncer de colo uterino e se as lesões iniciais forem tratadas de forma adequada, a redução da taxa de câncer cervical pode chegar a 90% (TAVARES & PRADO, 2006).

Desde 1988, no Brasil, o Ministério da Saúde (MS) tem seguido a recomendação da Organização Mundial de Saúde, que propõe um exame de Papanicolaou a cada três anos em mulheres entre 25 e 64 anos de idade, após dois controles anuais consecutivos com resultado normal. Em 1996, o MS, por meio do Instituto Nacional de Câncer (INCA), implementou o Programa Viva Mulher, envolvendo cinco capitais brasileiras e um estado, tendo como população-alvo mulheres pertencentes à faixa etária de 35 a 49 anos. Em 1998, as ações do Viva Mulher foram estendidas a todos os municípios

brasileiros por meio de uma campanha nacional para combater o câncer do colo de útero.

Entre os fatores para a não realização do exame, são apontados as barreiras de acesso aos serviços de saúde, escolaridade e nível socioeconômico das mulheres, questões culturais, medo e preconceito (JORGE *et al.*, 2012; MATOS *et al.*, 2011). Dada a importância e prioridade do exame Papanicolaou para a prevenção do câncer do colo de útero, este trabalho foi realizado. Diferente de outros estudos voltados para identificação dos fatores associados a não realização do exame Papanicolaou, o presente trabalho não se concentrou apenas na análise de mulheres em nível municipal (MARTINS *et al.*, 2006; AMORIM *et al.*, 2006; OLIVEIRA *et al.*, 2006), e sim em nível nacional. Outros estudos consideraram ainda um grupo menor de mulheres, como por exemplo: usuárias de alguma unidade de saúde; atendidas por uma Estratégia da Saúde da Família de um dado município (ANDRADE *et al.*, 2014); acadêmicas do curso de enfermagem de uma Universidade (CARVALHO *et al.*, 2011). Embora haja predominância de estudos focando na análise de mulheres na faixa de 25 a 59 anos de idade, há estudos que consideram outras faixas-etárias de interesse, inclusive a faixa etária de 25 anos ou mais, que foi a utilizada no presente estudo. No estudo de OLIVEIRA & KRETZMANN (2011), foram usados os dados da PNAD 2008 para identificar os fatores determinantes da realização do exame Papanicolaou em mulheres no Brasil, entretanto, ao considerar o método de máxima verossimilhança na modelagem estatística, os autores ignoraram as informações do plano amostral desta pesquisa. Cabe ressaltar que é necessário incorporar as informações do plano amostral ao ajustar modelos estatísticos, pois as estimativas pontuais dos parâmetros do modelo são influenciadas por pesos distintos das unidades da amostra, enquanto que as estimativas dos erros padrões são influenciadas também pela estrutura do plano amostral da pesquisa (estratificação e conglomeração) (MORAES *et al.*, 2012; SZWARCOWALD & DAMACENA, 2008).

2. Objetivos

2.1. Objetivo geral

Estudar a associação entre a não realização do exame preventivo para o câncer do colo do útero (Papanicolaou) e as características demográficas, socioeconômicas, comportamentais e de saúde relativas a mulheres de 25 anos ou mais de idade, no ano de 2008.

2.2. Objetivos específicos

- Estimar a prevalência da não realização de exame preventivo para câncer de colo do útero no Brasil.
- Descrever o perfil demográfico, socioeconômico, comportamental e de saúde das mulheres de 25 anos ou mais, considerando a realização e a não realização de exame preventivo para o câncer do colo do útero.
- Identificar a associação dos fatores demográficos, socioeconômicos, comportamentais e de saúde das mulheres de 25 anos ou mais, com a não realização do exame preventivo para o câncer do colo do útero.

3. Material e Métodos

3.1. Pesquisa Nacional por Amostra de Domicílios (PNAD)

A PNAD é um inquérito amostral de abrangência nacional, de periodicidade anual, realizado pelo IBGE. Em 2008 coletou informações de uma amostra probabilística de 150.591 domicílios e 391.868 pessoas (IBGE, 2009). A amostra foi planejada para obter estimativas representativas para o Brasil, Grandes Regiões, Unidades da Federação e nove Regiões Metropolitanas (Belém, Fortaleza, Recife, Salvador, Belo Horizonte, Rio de Janeiro, São Paulo, Curitiba e Porto Alegre).

Esta pesquisa investiga diversas características demográficas e socioeconômicas da população, que incluem características gerais, de educação, trabalho, rendimento e habitação, que estão permanentemente presentes nas pesquisas, bem como características sobre migração, fecundidade, nupcialidade, saúde, nutrição, entre outros. A periodicidade desta pesquisa varia conforme a necessidade de informação do País (IBGE, 2009). No ano de 2008, a PNAD, em seu suplemento, levantou dados sobre diferentes características de saúde da população brasileira relacionadas à necessidade de cuidados médicos e ao acesso e utilização de serviços de saúde, entre elas pode-se citar: a restrição de atividade por motivo de saúde, a incapacidade funcional, autoavaliação de saúde e a presença de doenças crônicas (IBGE, 2009).

3.1.1. Plano amostral da PNAD

A PNAD é realizada por meio de uma amostra probabilística de domicílios obtida em três estágios de seleção, onde as unidades primárias são os *municípios*; as unidades secundárias os *setores censitários*; e as unidades terciárias são as *unidades domiciliares* (IBGE, 2009).

No primeiro estágio de seleção, os municípios são classificados em dois grupos: autorrepresentativos (com probabilidade 1 de pertencer à amostra) e não autorrepresentativos; sendo que os municípios não autorrepresentativos passam por um processo de estratificação e, para cada estrato formado, eles são selecionados com reposição e com probabilidade proporcional à população residente obtida no Censo Demográfico 2000.

No segundo estágio, os setores censitários, são selecionados, por sua vez, em cada município da amostra com reposição e com probabilidade proporcional ao número de unidades domiciliares existentes por ocasião do Censo Demográfico 2000.

Finalmente, no terceiro e último estágio, as unidades domiciliares são selecionadas, com equiprobabilidade, em cada setor censitário da amostra. Nos domicílios selecionados, são realizadas entrevistas para levantamento das características habitacionais e dos seus moradores. Não há seleção de moradores, isto é, todos os moradores dos domicílios da amostra são incluídos na pesquisa.

Na seleção das unidades primárias e secundárias (municípios e setores censitários) da PNAD foram adotadas a divisão territorial e a malha setorial vigentes em 1º de agosto de 2000 e utilizadas para a realização do Censo Demográfico 2000 (IBGE, 2009). Além disso, antes da seleção de domicílios, realizou-se, em todos os setores da amostra, a operação de listagem de todas as unidades residenciais e não residenciais existentes na área visando obter um cadastro básico atualizado das unidades domiciliares.

Do exposto, observa-se que a PNAD 2008 é uma pesquisa que utiliza um plano amostral complexo que combina estratificação, conglomeração e probabilidades desiguais de seleção. Portanto, a análise estatística de dados de pesquisas amostrais complexas *“só se torna possível se forem incluídas no arquivo de dados variáveis que informem a estrutura do plano amostral, identificando ao menos o estrato, a UPA e o peso de cada unidade da amostra* (PESSOA e SILVA, 1998:156)”, uma vez que as estimativas pontuais dos parâmetros de interesse são influenciadas por pesos distintos das observações e as estimativas de variância são influenciadas pela conglomeração, estratificação e pesos amostrais (SZWARCOWALD *et al.*, 2008; PESSOA *et al.*, 2002; MORAES *et al.*, 2012). Além da obtenção de estimativas pontuais corretas, há também necessidade de obtenção de estimativas de variância, pois são medidas utilizadas na avaliação da precisão das estimativas pontuais, sendo necessárias para a construção de intervalos de confiança e para a realização de testes de hipóteses para os coeficientes de modelos estatísticos. Desse modo, segundo ASPAROUHOV (2005) se os conglomerados não forem levados em consideração na análise dos dados espera-se uma subestimação das variâncias, enquanto que com a não consideração da estratificação espera-se uma sobrestimação das variâncias.

No banco de dados da PNAD 2008, estão disponíveis informações sobre os pesos amostrais associados aos indivíduos da amostra, além das informações estruturais

(estratos e unidades primárias de amostragem) do plano amostral adotado nesta pesquisa. Para incorporação do plano amostral da PNAD 2008, é necessário utilizar pacotes estatísticos que permitem incorporar tais informações na análise descritiva e na modelagem estatística. Neste trabalho foi utilizado o software SPSS for Windows, versão 20.0 (SPSS Inc., Chicago, Estados Unidos), para ajustar o modelo logístico binário para a não realização de exame preventivo para câncer do colo do útero, considerando os pesos amostrais e as informações estruturais do plano amostral da PNAD 2008.

3.2. População de estudo

A população de estudo é composta por mulheres de 25 anos ou mais de idade que não autorreferiram câncer, residentes em domicílios particulares permanentes⁴ no Brasil, em 2008. Dentre essas mulheres excluiu-se aquelas que não relataram informações (missings) sobre pelo menos uma das variáveis explicativas consideradas no estudo.⁵

3.3. Variáveis utilizadas no estudo

As variáveis utilizadas no estudo (vinte), apresentadas no Quadro 1, se referem a características socioeconômicas, demográficas, comportamentais e de saúde e outras, de mulheres de 25 anos ou mais. A variável resposta, relacionada à realização do exame preventivo de câncer de colo do útero, é de natureza dicotômica com as seguintes categorias: 1) *Não*, se a mulher “nunca fez” exame preventivo para câncer do colo do útero; 2) *Sim*, se a mulher já fez este tipo de exame.

⁴ Domicílios destinados à moradia de uma pessoa ou de um grupo de pessoas cujo relacionamento é ditado por laços de parentesco, dependência doméstica, ou normas de convivência.

⁵ Tamanho da amostra 60.483 mulheres (registros).

Quadro 1: Variáveis demográficas, socioeconômicas, comportamentais e de saúde relativas a mulheres de 25 anos ou mais de idade.

<i>Variáveis de Estudo</i>	<i>Categorias</i>
<i>Características Demográficas</i>	
Faixa Etária	De 25 a 29 anos, de 30 a 39 anos, de 40 a 49 anos, de 50 a 59 anos, 60 a 69anos, 70 anos ou mais.
Cor/Raça	Branca, Não Branca.
Região de Residência	Norte, Nordeste, Sudeste, Sul, Centro-Oeste.
Área de Localização do Domicílio	Urbana, Rural.
<i>Características Socioeconômicas</i>	
Anos de Estudo	Sem instrução e menos de 1 ano, 1 a 3 anos, 4 a 7 anos, 8 a 10 anos, 11 anos ou mais.
Situação de Ocupação da Mulher	Ocupada, Desocupada.
Renda Domiciliar Mensal <i>per capita</i>	Sem renda ou até 1 s.m,mais de 1 a 5 s.m, mais de 5 s.m.
<i>Características Comportamentais</i>	
Atividade Física	Pratica, não pratica.
Tabagismo	Fumante, ex-fumante, não fumante.
<i>Características de Saúde</i>	
Exame Preventivo para Câncer de Colo do Útero (variável resposta)	Realizado, Não realizado.
Morbidade Autorreferida (com exceção de câncer e depressão)	Nenhuma doença crônica, 1 doença crônica, 2 doenças crônicas, 3 ou mais doenças crônicas.
Depressão	Sim, Não.
Autoavaliação de Saúde	Muito Bom ou Bom, Regular, Ruim ou Muito Ruim.
Posse de Plano de Saúde	Sim, Não.
Consulta Médica nos Últimos 12 Meses	Sim, Não.
Utilização de Medicamento de Uso contínuo (MUC).	Sim, Não.

Limitação ⁶	Tem dificuldade (pequena), não tem dificuldade.
Mamografia	Sim, Não.
Outras Características	
Tem Mãe Viva	Sim, Não.
Teve Filho	Sim, Não.

3.4. Modelo de regressão logística binária

3.4.1. Especificação do modelo

O modelo de regressão logística binária é um modelo linear generalizado onde a variável resposta tem distribuição de probabilidade de Bernoulli e a função de ligação é a função logística, que pode ser representada pela equação abaixo:

$$\ln \left(\frac{p_{abcdefghijklmnopqrs}}{1 - p_{abcdefghijklmnopqrs}} \right) = \eta_{abcdefghijklmnopqrs} ,$$

onde o preditor linear $\eta_{abcdefghijklmnopqrs}$ do modelo é dado por:

$$\begin{aligned} \eta_{abcdefghijklmnopqrs} = & \mu + \alpha_a + \zeta_b + \gamma_c + \lambda_d + \delta_e + \psi_f + \theta_g + \Phi_h + \pi_i + \rho_j + \\ & + \nu_k + \tau_l + \varphi_m + \xi_n + \omega_o + \varpi_p + \kappa_q + \varsigma_r + E_s \end{aligned}$$

Assim, a probabilidade da mulher não realizar o exame preventivo de câncer do colo do útero é dada por:

$$p_{abcdefghijklmnopqrs} = \frac{e^{\eta_{abcdefghijklmnopqrs}}}{1 + e^{\eta_{abcdefghijklmnopqrs}}}$$

⁶Limitação na realização das atividades básicas do cotidiano, como alimentar-se, tomar banho e ir ao banheiro. As mulheres com “grande dificuldade” e “impossibilidade” na realização dessas atividades não permaneceram na amostra após a exclusão dos missings.

sendo que,

$\mu \rightarrow$ intercepto do modelo.

$\alpha_a \rightarrow$ efeito do a -ésimo nível do fator *faixa etária*; $a=1, 2, 3, 4, 5, 6$.

$\zeta_b \rightarrow$ efeito do b -ésimo nível do fator *cor/raça*; $b=1, 2$.

$\gamma_c \rightarrow$ efeito do c -ésimo nível do fator *região residência*; $c=1, 2, 3, 4, 5$.

$\lambda_d \rightarrow$ efeito do d -ésimo nível do fator de *área de localização do domicílio*; $d=1,$

2.

$\delta_e \rightarrow$ efeito do e -ésimo nível do fator *anos de estudo*; $e=1, 2, 3, 4, 5$.

$\psi_f \rightarrow$ efeito do f -ésimo nível do fator *situação de ocupação da mulher*; $f=1, 2$.

$\theta_g \rightarrow$ efeito do g -ésimo nível do fator *renda domiciliar mensal per capita*; $g=1,$

2, 3.

$\phi_h \rightarrow$ efeito do h -ésimo nível do fator *atividade física*; $h=1, 2$.

$\pi_i \rightarrow$ efeito do i -ésimo nível do fator *tabagismo*; $i=1, 2, 3$.

$\rho_j \rightarrow$ efeito do j -ésimo nível do fator *morbidade autoreferida*; $j=1, 2, 3, 4$.

$\nu_k \rightarrow$ efeito do k -ésimo nível do fator *depressão*; $k=1, 2$.

$\tau_l \rightarrow$ efeito do l -ésimo nível do fator *autoavaliação de saúde*; $l=1, 2, 3$.

$\varphi_m \rightarrow$ efeito do m -ésimo nível do fator *posse de plano de saúde*; $m=1, 2$.

$\xi_n \rightarrow$ efeito do n -ésimo nível do fator *consulta médica nos últimos 12 meses*;

$n=1, 2$.

$\omega_o \rightarrow$ efeito do o -ésimo nível do fator *utilização de medicamento de uso contínuo (MUC)*; $o=1, 2$.

$\varpi_p \rightarrow$ efeito do p -ésimo nível do fator *limitação*; $p=1, 2$.

$\kappa_q \rightarrow$ efeito do q -ésimo nível do fator *mamografia*; $q=1, 2$.

$\varsigma_r \rightarrow$ efeito do r -ésimo nível do fator *tem mãe viva*; $r=1, 2$.

$E_s \rightarrow$ efeito do s -ésimo nível do fator *teve filho*; $s=1, 2$.

3.4.2. Métodos de estimação dos parâmetros do modelo

3.4.2.1. Método de máxima verossimilhança (MV)

Se $y_1, y_2, \dots, y_n$ são valores observados das variáveis $Y_1, Y_2, \dots, Y_n$, independentes, e todas com distribuição de Bernoulli com parâmetro p_i , $i=1, \dots, n$, a função de densidade de probabilidade de Y_i é dada por:

$$P(Y_i = y_i | p_i) = p_i^{y_i} (1 - p_i)^{1-y_i}; i=1, 2, \dots, n,$$

onde,

$$p_i = p_i(\beta) = \frac{e^{x_i' \beta}}{1 + e^{x_i' \beta}}$$

sendo que,

$$\beta = [\mu, \alpha_1, \dots, \alpha_5, \zeta_1, \gamma_1, \dots, \gamma_4, \lambda_1, \delta_1, \dots, \delta_4, \psi_1, \theta_1, \theta_2, \Phi_1, \pi_1, \pi_2, \rho_1, \dots, \rho_3, \nu_1, \tau_1, \tau_2, \varphi_1, \xi_1, \omega_1, \varpi_1, \kappa_1, \varsigma_1, \varepsilon_1].$$

A função de verossimilhança da amostra é dada por:

$$L(\beta | y) = \prod_{i=1}^n P(Y_i = y_i | p_i) = \prod_{i=1}^n p_i^{y_i} (1 - p_i)^{1-y_i}.$$

A estimativa $\hat{\beta}$ obtida pelo método de MV é o valor de β que maximiza a função de verossimilhança $L(\beta | y)$.

Como maximizar uma função é equivalente a maximizar o seu logaritmo, aplicou-se o logaritmo à função de verossimilhança da amostra, logo:

$$\begin{aligned} \ln L(\beta | y) &= \ln \prod_{i=1}^n P(Y_i = y_i | p_i) = \sum_{i=1}^n \ln P(Y_i = y_i | p_i) \\ &= \sum_{i=1}^n [y_i \ln(p_i) + (1 - y_i) \ln(1 - p_i)] \\ &= \sum_{i=1}^n [y_i \ln(p_i) + \ln(1 - p_i) - y_i \ln(1 - p_i)] \\ &= \sum_{i=1}^n \left[y_i \ln \left(\frac{p_i}{1 - p_i} \right) + \ln(1 - p_i) \right] \end{aligned}$$

$$= \sum_{i=1}^n [y_i x_i' \beta - \ln(1 + e^{x_i' \beta})]$$

Derivando $\ln L(\beta | y)$ em relação a cada um dos parâmetros do modelo, obtém-se:

$$\begin{aligned} \frac{\partial \ln L(\beta | y)}{\partial \beta} &= \sum_{i=1}^n \frac{\partial}{\partial \beta} [y_i x_i' \beta - \ln(1 + e^{x_i' \beta})] \\ &= \sum_{i=1}^n u_i(\beta), \end{aligned}$$

onde,

$$u_i(\beta) = \frac{\partial}{\partial \beta} [y_i x_i' \beta - \ln(1 + e^{x_i' \beta})]$$

é o vetor dos escores do indivíduo i de dimensão $p \times 1$.

Ao igualar as derivadas parciais à zero, obtém-se o seguinte sistema de equações não lineares:

$$\frac{\partial \ln L(\beta | y)}{\partial \beta} = \sum_{i=1}^n u_i(\beta) = 0.$$

A partir desse sistema de equações, calcula-se a estimativa de MV de β , denotada por $\hat{\beta}_{MV}$. Desse modo, a probabilidade estimada em favor da ocorrência do evento de interesse referente ao i -ésimo indivíduo é dada por:

$$\hat{p}_i = \hat{p}_i(\hat{\beta}_{MV}) = \frac{e^{x_i' \hat{\beta}_{MV}}}{1 + e^{x_i' \hat{\beta}_{MV}}},$$

onde,

$$\begin{aligned} \hat{\beta}_{MV} &= [\hat{\mu}, \hat{\alpha}_1, \dots, \hat{\alpha}_5, \hat{\xi}_1, \hat{\gamma}_1, \dots, \hat{\gamma}_4, \hat{\lambda}_1, \hat{\delta}_1, \dots, \hat{\delta}_4, \hat{\psi}_1, \hat{\theta}_1, \hat{\theta}_2, \hat{\Phi}_1, \hat{\pi}_1, \hat{\pi}_2, \hat{\rho}_1, \dots, \hat{\rho}_3, \\ &\hat{\nu}_1, \hat{\tau}_1, \hat{\tau}_2, \hat{\phi}_1, \hat{\xi}_1, \hat{\omega}_1, \hat{\omega}_1, \hat{\kappa}_1, \hat{\zeta}_1, \hat{E}_1]. \end{aligned}$$

Para grandes amostras ($n \rightarrow \infty$):

$$VAR(\hat{\beta}_{MV}) = -\mathcal{J}^{-1}$$

$$\widehat{VAR}(\hat{\beta}_{MV}) = -\hat{\mathcal{J}}^{-1} = -\hat{\mathcal{J}}^{-1}(\hat{\beta}_{MV}),$$

sendo que,

$$\hat{\mathfrak{T}} = \sum_{i=1}^n \frac{\partial}{\partial \beta} u_i(\beta) \Big|_{\beta=\hat{\beta}_{MV}}.$$

3.4.2.2. Método de máxima pseudo-verossimilhança (MPV)

A amostra da PNAD é selecionada utilizando um plano amostral complexo, portanto foi necessário adotar outro método de estimação que permite incorporar tanto os pesos amostrais quanto as informações estruturais (estratificação e conglomeração) desse plano amostral. Assim, os parâmetros do modelo de regressão logística ajustada a partir dos dados da PNAD são estimados empregando o método de máxima pseudo-verossimilhança (MPV).

A estimação pelo método de MPV é baseada no princípio da modelagem de superpopulação, no qual cada unidade populacional i é associada a uma variável aleatória Y_i , $i=1,2,\dots, N$ (onde N é o tamanho da população), para a qual se especifica a forma de sua distribuição de probabilidade representada por um modelo denominado “modelo de superpopulação”, que é usado no processo de inferência sobre os parâmetros.

O modelo de superpopulação é uma função que representa a distribuição conjunta de $\mathbf{Y}=(Y_1, Y_2,\dots, Y_N)$, denotada por $P(\mathbf{Y}=\mathbf{y}|\boldsymbol{\beta})$, onde $\mathbf{y}=(y_1, y_2,\dots, y_N)$ é um vetor de observações da variável resposta $\mathbf{Y}$ na população, e $\boldsymbol{\beta}$ é o vetor de parâmetros de interesse cuja dimensão é $p \times 1$. Desse modo, o modelo de superpopulação resultante é dado pelo produto da função de densidade de cada variável aleatória Y_i , isto é:

$$P(\mathbf{Y} = \mathbf{y}|\boldsymbol{\beta}) = \prod_{i=1}^N p_i^{y_i} (1 - p_i)^{1-y_i}$$

Sendo $y_1, y_2,\dots, y_N$ observações das variáveis aleatórias $Y_1, Y_2,\dots, Y_N$ independentes, todas com distribuição de Bernoulli com parâmetro p_i , a função de probabilidade de Y_i é dada por:

$$P(Y_i = y_i|p_i) = p_i^{y_i}(1 - p_i)^{1-y_i}; i=1, 2, \dots, N,$$

onde

$$p_i = p_i(\boldsymbol{\beta}) = \frac{e^{\mathbf{x}_i' \boldsymbol{\beta}}}{1 + e^{\mathbf{x}_i' \boldsymbol{\beta}}},$$

sendo que,

$$\boldsymbol{\beta} = [\mu, \alpha_1, \dots, \alpha_5, \zeta_1, \gamma_1, \dots, \gamma_4, \lambda_1, \delta_1, \dots, \delta_4, \psi_1, \theta_1, \theta_2, \Phi_1, \pi_1, \pi_2, \rho_1, \dots, \rho_3, \nu_1, \tau_1, \tau_2, \varphi_1, \xi_1, \omega_1, \varpi_1, \kappa_1, \varsigma_1, E_1].$$

A estimação de MPV consiste em maximizar a função de verossimilhança populacional, dada por:

$$L(\boldsymbol{\beta} | \mathbf{y}) = \prod_{i=1}^N P(Y_i = y_i, p_i) = \prod_{i=1}^N p_i^{y_i} (1 - p_i)^{1-y_i}$$

O logaritmo da função de verossimilhança populacional é dado por:

$$\ln L(\boldsymbol{\beta} | \mathbf{y}) = \ln \prod_{i=1}^N P(Y_i = y_i | p_i) = \sum_{i=1}^N \ln P(Y_i = y_i | p_i)$$

Derivando $\ln L(\boldsymbol{\beta} | \mathbf{y})$ em relação a cada um dos parâmetros do modelo e igualando as derivadas parciais à zero, obtém-se um sistema de equações de verossimilhanças populacionais dadas por:

$$\frac{\partial \ln L(\mathbf{p} | \mathbf{y})}{\partial \boldsymbol{\beta}} = \sum_{i=1}^N \mathbf{u}_i(\boldsymbol{\beta}) = \mathbf{0},$$

onde,

$$\mathbf{u}_i(\boldsymbol{\beta}) = \frac{\partial}{\partial \boldsymbol{\beta}} [y_i \mathbf{x}_i' \boldsymbol{\beta} - \ln(1 + e^{\mathbf{x}_i' \boldsymbol{\beta}})]$$

é o vetor dos escores do indivíduo i de dimensão $p \times 1$.

Como $\sum_{i=1}^N \mathbf{u}_i(\boldsymbol{\beta}) = \mathbf{T}(\boldsymbol{\beta})$ é a soma dos vetores de escores na população, para estimar este vetor de totais populacionais pode utilizar o estimador linear ponderado dado por: $\hat{\mathbf{T}} = \sum_{i=1}^n w_i \mathbf{u}_i(\boldsymbol{\beta})$, onde w_i são os pesos associados as unidades da amostra.

Fazendo $\hat{\mathbf{T}} = \mathbf{0}$ obtém-se um sistema de equações de pseudo-verossimilhança, cuja solução é o estimador de MPV do pseudo-parâmetro (PP), sendo denotado por $\boldsymbol{\beta}_{PP}$ (e consequentemente de $\boldsymbol{\beta}$). Como foi realizada uma pesquisa por amostragem e não censitária, encontra-se disponível apenas uma amostra da população de interesse, selecionada utilizando o plano amostral apresentado na seção 3.1.1. Desse modo, não é

possível calcular os pseudo-parâmetros, restando como alternativa adequada determinar estimadores consistentes para β_{PP} . Essas quantidades β_{PP} desempenham um papel fundamental nessa abordagem inferencial, por se constituírem em pseudo-parâmetros, eleitos como alvo da inferência em um esquema, que incorpora o planejamento amostral (PESSOA & SILVA, 1998). Além disso, como em pesquisas por amostragem o tamanho da população (N) é geralmente grande, um estimador adequado para o pseudo-parâmetro, geralmente, também é adequado para o parâmetro do modelo de superpopulação (PESSOA & SILVA, 1998).

A PNAD utilizada nesse estudo é uma pesquisa de abrangência nacional, portanto o vetor de estimativas de MPV, denotado por $\hat{\beta}_{MPV}$, obtido para o vetor de pseudo-parâmetros β_{PP} também serão válidas para o vetor de parâmetros de interesse β . Desse modo, a probabilidade estimada em favor da ocorrência do evento de interesse referente ao i -ésimo indivíduo é dada por:

$$\hat{p}_i = \hat{p}_i(\hat{\beta}_{MPV}) = \frac{e^{x_i' \hat{\beta}_{MPV}}}{1 + e^{x_i' \hat{\beta}_{MPV}}}.$$

A estimação da variância assintótica (PESSOA & SILVA, 1988) dos estimadores de MPV dos parâmetros do modelo é feita empregando-se a seguinte expressão:

$$\begin{aligned} \widehat{VAR}(\hat{\beta}_{MPV}) &\cong [\mathcal{J}(\hat{\beta}_{MPV})]^{-1} \widehat{VAR}[\hat{T}(\hat{\beta}_{MPV})] [\mathcal{J}(\hat{\beta}_{MPV})]^{-1} \\ \widehat{VAR}(\hat{\beta}_{MPV}) &\cong [\mathcal{J}(\hat{\beta}_{MPV})]^{-1} \widehat{VAR} \left[\sum_{i=1}^n w_i u_i(\hat{\beta}_{MPV}) \right] [\mathcal{J}(\hat{\beta}_{MPV})]^{-1}, \end{aligned}$$

onde,

$$\mathcal{J}(\hat{\beta}_{MPV}) = \left. \frac{\partial \hat{T}(\beta)}{\partial \beta} \right|_{\beta = \hat{\beta}_{MPV}} = \sum_{i=1}^n w_i \left. \frac{\partial u_i(\beta)}{\partial \beta} \right|_{\beta = \hat{\beta}_{MPV}}$$

Segundo BINDER (1983), o vetor de estimadores $\hat{\beta}_{MPV}$ tem distribuição assintoticamente normal multivariada $\hat{\beta}_{MPV} \sim N[\beta, \widehat{VAR}(\hat{\beta}_{MPV})]$, o que torna válida a realização de inferências para β no caso de amostras suficientemente grandes. Alternativamente, utiliza-se a seguinte estatística:

$$[\widehat{VAR}(\hat{\beta}_{MPV})]^{-\frac{1}{2}} (\hat{\beta}_{MPV} - \beta) \sim N(0, I)$$

3.4.3. Inferência no modelo

3.4.3.1. Teste de Wald de significância geral dos parâmetros do modelo

Seja Para verificar a significância dos parâmetros do modelo será utilizado o teste de Wald de significância geral, que é descrito a seguir:

- **Hipóteses a serem testadas:**

$$\begin{cases} H_0: C\beta = c \\ H_1: C\beta \neq c \end{cases}$$

Onde,

$\beta = (\beta_1, \dots, \beta_P)'$ é o vetor de parâmetros desconhecidos, de dimensão $P \times 1$, do modelo de regressão.

C é uma matriz de dimensão $R \times P$ de posto completo $R=P-Q$; e c é um vetor $R \times 1$.

- **Estatística de teste:**

Sob $H_0: C\beta = c$, a estatística de teste é dada por:

$$X_W^2 = (C\hat{\beta}_{MPV} - c)[C\widehat{VAR}(\hat{\beta}_{MPV})C']^{-1}(C\hat{\beta}_{MPV} - c),$$

onde $\hat{\beta}_{MPV}$ é o estimador de MPV do vetor β de parâmetros do modelo e $\widehat{VAR}(\hat{\beta}_{MPV})$ é o estimador da matriz de variância-covariância do estimador de MPV do vetor β de parâmetro do modelo.

Segundo BINDER (1983), a normalidade assintótica do estimador $\hat{\beta}_{MPV}$ e a consistência do estimador da matriz de variância-covariância de $\hat{\beta}_{MPV}$, permite afirmar que:

$$X_W^2 \sim X_R^2$$

Com o objetivo de melhorar a aproximação de estatística de teste, pode-se utilizar a estatística *F-corrigida de Wald*, que está implementada em pacotes estatísticos para a análise de dados provenientes de pesquisas amostrais complexas, dada por:

$$F = \frac{X_W^2}{R} \sim F(R, v),$$

onde $v=m-H$, sendo m o número de UPAs e H o número de estratos da amostra, indicando o número de graus de liberdade.

- **Região crítica (RC):**

$$RC = \{f \in R \mid f \geq f_{\alpha, R, v}\}.$$

- **Tomada de decisão:**

Se o valor observado f_{obs} da estatística de teste F pertencer a região crítica (RC), rejeita-se $H_0: \mathbf{C}\boldsymbol{\beta} = \mathbf{c}$ ao nível de significância α . Equivalentemente, rejeita-se H_0 se o p-valor $\leq \alpha$.

3.4.3.2. Teste de Wald de significância individual dos parâmetros do modelo

Para verificar a significância individual dos parâmetros do modelo será utilizado o teste de Wald, descrito a seguir:

- **Hipóteses a serem testadas:**

$$\begin{cases} H_0: \beta_j = 0 \\ H_1: \beta_j \neq 0 \end{cases},$$

onde β_j é o efeito principal do j -ésimo nível de um determinado fator (F); $j=1, \dots, f$.

- **Estatística de teste:**

$$\text{Sob } H_0: \beta_j = 0, \text{ temos que } Z = \frac{\hat{\beta}_j}{\sqrt{\widehat{VAR}(\hat{\beta}_j)}} \stackrel{a}{\sim} N(0,1).$$

Sendo que, $\hat{\beta}_j$ é o estimador de MPV do parâmetro β_j do modelo e $\widehat{VAR}(\hat{\beta}_j)$ é o estimador da variância do estimador de MPV do parâmetro β_j .

- **Região crítica (RC):**

$$RC = \{z \in R \mid z < -z_{\frac{\alpha}{2}} \text{ ou } z > z_{\frac{\alpha}{2}}\}$$

- **Tomada de decisão:**

Se o valor observado da estatística de teste (z_{obs}) pertence à região crítica (RC) rejeita-se $H_0: \beta_j = 0$ ao nível de significância α , ou seja, existe efeito estatisticamente

significante do j -ésimo nível do fator F . Equivalentemente, rejeita-se H_0 se o p-valor $\leq \alpha$.

3.4.4. Medidas de qualidade do ajuste

Na regressão logística é comum usar as seguintes medidas de pseudo- R^2 para avaliar a qualidade do ajuste do modelo que são baseadas na comparação do logaritmo da função de pseudo-verossimilhança do modelo final ajustado (*modelo selecionado*), denotado por $\ln L(\hat{\beta}_1 | \mathbf{y})$, com o logaritmo da função de pseudo-verossimilhança do modelo composto apenas pelo intercepto (*modelo nulo*), denotado por $\ln L(\hat{\beta}_0 | \mathbf{y})$. Neste trabalho foram utilizados os *pseudo- R^2 de Nagelkerke* (R_N^2) e de *Mc Fadden* (R_{MF}^2).

O *Pseudo- R^2 de Nagelkerke* é uma correção do *Pseudo- R^2 de Cox & Snell*, realizada para que esta medida possa assumir o valor 1 no caso do ajuste ser perfeito, estando no intervalo $[0, 1]$. Esta medida é calculada por:

$$R_N^2 = \frac{R_{CS}^2}{\frac{2[\ln L(\hat{\beta}_0 | \mathbf{y})]}{1 - e^{-\frac{2}{\sum_{i=1}^n w_i}}}}; \quad 0 \leq R_N^2 \leq 1, \text{ onde:}$$

$$R_{CS}^2 = 1 - e^{-\frac{2[\ln L(\hat{\beta}_0 | \mathbf{y}) - \ln L(\hat{\beta}_1 | \mathbf{y})]}{\sum_{i=1}^n w_i}}; \quad 0 \leq R_{CS}^2 < 1.$$

O *pseudo- R^2 de McFadden* (R_{MF}^2) é dado por:

$$R_{MF}^2 = 1 - \frac{\ln L(\hat{\beta}_1 | \mathbf{y})}{\ln L(\hat{\beta}_0 | \mathbf{y})} \quad \text{ou} \quad R_{MF}^2 = \frac{\ln L(\hat{\beta}_0 | \mathbf{y}) - \ln L(\hat{\beta}_1 | \mathbf{y})}{\ln L(\hat{\beta}_0 | \mathbf{y})}; \quad \text{onde } 0 \leq R_{MF}^2 < 1,$$

Essas medidas indicam o ganho de informação obtido pelo modelo final em comparação ao modelo nulo (sem as variáveis explicativas).

3.4.5. Avaliação da capacidade preditiva do modelo

3.4.5.1. Taxa de classificações corretas

Com relação à avaliação da capacidade preditiva do modelo será utilizada a estratégia de comparar as categorias observadas e preditas pelo modelo selecionado. A seguir é representada a tabela de contingência (Quadro 2) que possibilita comparar tais categorias.

Quadro 2: Distribuição dos indivíduos segundo as categorias observadas e preditas pelo modelo de regressão logística binária.

Categorias Preditas	Categorias observadas		Total
	Sucesso (Y=1)	Fracasso (Y=0)	
Sucesso ($\hat{Y}=1$)	o_{11} (Verdadeiros Positivos)	o_{12} (Falsos Positivos)	$o_{11} + o_{12}$
Fracasso ($\hat{Y}=0$)	o_{21} (Falsos Negativos)	o_{22} (Verdadeiros Negativos)	$o_{21} + o_{22}$
Total	$o_{11} + o_{21}$	$o_{12} + o_{22}$	$o_{11} + o_{12} + o_{21} + o_{22}$

A capacidade preditiva (ou classificatória) do modelo pode também ser avaliada pela sensibilidade e especificidade do modelo, além da taxa global de classificações corretas.

A *sensibilidade* (S) é a proporção de verdadeiros positivos entre todas as mulheres que não realizaram o exame preventivo do câncer do colo de útero. Em outras palavras, a sensibilidade avalia a capacidade do modelo classificar a mulher como tendo não realizado o exame, quando de fato este exame não foi realizado pela mulher.

$$S = P(\hat{Y} = 1 | Y = 1) = \frac{o_{11}}{o_{11} + o_{21}}$$

A *especificidade* (E) é a proporção de verdadeiros negativos entre todos as mulheres que realizaram o exame preventivo do câncer do colo do útero. Em outras palavras, a especificidade avalia a capacidade do modelo classificar a mulher como tendo realizado o exame preventivo, quando de fato este foi realizado pela mulher.

$$E = P(\hat{Y} = 0|Y = 0) = \frac{o_{22}}{o_{12} + o_{22}}$$

A taxa global (%) de classificações corretas, por sua vez, é definida pela razão entre o número de mulheres classificadas corretamente pelo modelo e o número total de mulheres consideradas no estudo.

$$G = \frac{o_{11} + o_{22}}{o_{11} + o_{21} + o_{12} + o_{22}}$$

3.4.5.2 Área sob a curva ROC

A curva ROC (*Receiver Operating Characteristic*) é um gráfico que permite estudar a variação da sensibilidade e especificidade para diferentes pontos de corte. No eixo das ordenadas representa-se os valores de sensibilidade (proporção de verdadeiros positivos) e no eixo das abscissas os valores dos complementos de especificidade (proporção de falsos positivos).

Na seção 3.4.5.1 deste trabalho já foi apresentando o conceito de *sensibilidade* (S) e *especificidade* (E), sendo representadas por:

$$S = P(\hat{Y} = 1|Y = 1) = \frac{o_{11}}{o_{11} + o_{21}}$$

$$E = P(\hat{Y} = 0|Y = 0) = \frac{o_{22}}{o_{12} + o_{22}}$$

O complemento da especificidade, por sua vez, é a proporção de falsos positivos, que indica a capacidade do modelo classificar incorretamente a mulher como não tendo realizado o exame Papanicolaou, quando na realidade a mulher realizou o exame.

$$\bar{E} = 1 - E = 1 - P(\hat{Y} = 0|Y = 0) = P(\hat{Y} = 1|Y = 0) = \frac{o_{12}}{o_{12} + o_{22}}$$

O Quadro 2 (seção 3.4.5.1) apresenta os possíveis resultados que podem ocorrer no ajuste de um modelo de regressão logística binária explicativo da não realização do exame preventivo para câncer do colo de útero.

A área sob a curva ROC, denotada por A , que varia no intervalo compreendido entre 0 e 1, é uma outra medida de capacidade do modelo para discriminar a mulher como “não realizou o exame Papanicolaou” e “que realizou o exame Papanicolaou”.

Se a área for igual a 0,5 o modelo não consegue discriminar os indivíduos “que não realizaram o exame preventivo para câncer do colo do útero” e “que realizaram o exame preventivo para câncer do colo do útero”. Por outro lado, quanto mais próxima a curva estiver do canto superior esquerdo do gráfico, ou seja, maior a área sob a curva ROC ($A \geq 0,5$), e portanto maior a capacidade discriminatória do modelo. Segundo HOSMER & LEMESHOW (2000) o poder discriminatório do modelo pode ser classificado segundo os valores da área sob a curva ROC como mostrado no quadro 3:

Quadro 3: Classificação do poder discriminatório do modelo logístico binário baseada na área sob a curva ROC.

Área sob a curva ROC	Poder discriminatório do modelo
0,5	Sem poder discriminatório
$0,5 \leq A < 0,70$	Discriminação ruim
$0,70 \leq A < 0,80$	Discriminação aceitável
$0,80 \leq A < 0,90$	Discriminação boa
$A \geq 0,90$	Discriminação ótima

4. Resultados e Discussões

4.1 Resultados

Na Tabela 1 é apresentado o percentual de mulheres com 25 anos ou mais de idade segundo as suas características demográficas e socioeconômicas. Além disso, para cada uma dessas características, a Tabela 1, fornece o percentual de mulheres de 25 anos ou mais segundo a realização ou não do exame Papanicolaou.

Entre as mulheres de 25 anos ou mais de idade, observa-se um maior percentual de mulheres na faixa etária de *30 a 39 anos* (31,8%); residindo na *região Sudeste* (43,4%); com *11 anos ou mais de estudo* (46,1%); e com renda domiciliar *per capita* de no máximo 1 salário mínimo (49,0%). Verificou-se ainda que: 50,6% das mulheres que autorreferiram ser de raça *branca*; 85,0% moram na área *urbana*; e que 94,6% *são ocupadas* (Tabela 1).

As menores porcentagens de não realização do exame preventivo para câncer do colo de útero (exame Papanicolaou) se encontra nos grupos de mulheres com idades de *30 a 39 anos* (11,4%), *40 a 49 anos* (9,0%) e de *50 a 59 anos* (10,2%). O percentual de mulheres brancas que não realizaram o exame Papanicolaou (10,8%) é menor que o das mulheres não brancas (15,2%). Com relação as *grandes regiões brasileiras*, observou-se que as mulheres residentes nas regiões Sudeste e Sul são as que mais fizeram o exame Papanicolaou (isto é, apresentam menores percentuais de não realização do exame: 10,4% e 10,9%, respectivamente). Na área urbana, existe menor percentual de mulheres que não realizaram o exame (Urbana: 11,3% vs Rural: 22,3%)(Tabela 1).

Com relação a *escolaridade* e a *renda domiciliar mensal per capita*, observou-se que quanto maior é o tempo de estudo das mulheres e maior a renda domiciliar *per capita*, menor a porcentagem de mulheres que não realizaram o exame Papanicolaou. Quanto a *situação de ocupação*, verificou-se que as mulheres desocupadas são as que menos realizam o exame (ocupada: 12,9% vs desocupada: 14,7%).

Tabela 1: Distribuição percentual de mulheres com 25 anos ou mais de idade, segundo características demográficas e socioeconômicas no Brasil, e pelo percentual de condição de realização do exame preventivo para câncer de colo do útero (Papanicolaou).

Características	Mulheres com 25 anos ou mais	Exame Papanicolaou		Total
		Realizou o exame	Não realizou o exame	
Faixa Etária				
25 a 29 anos	18,4	79,6	20,4	100
30 a 39 anos	31,8	88,6	11,4	100
40 a 49 anos	27,2	91,0	9,0	100
50 a 59 anos	15,7	89,8	10,2	100
60 a 69 anos	5,4	81,4	18,6	100
70 anos ou mais	1,5	65,8	34,2	100
Cor / Raça				
Não Branca	49,4	84,8	15,2	100
Branca	50,6	89,2	10,8	100
Região de Residência				
Norte	6,7	86,4	13,6	100
Nordeste	26,4	81,8	18,2	100
Sudeste	43,4	89,6	10,4	100
Sul	16,0	89,1	10,9	100
Centro-Oeste	7,4	87,3	12,7	100
Área de Localização de Domicílio				
Urbana	85,0	88,7	11,3	100
Rural	15,0	77,7	22,3	100
Anos de Estudo				
Sem instrução e menos de 1 ano	8,2	70,5	29,5	100
1 a 3 anos	8,9	82,3	17,7	100
4 a 7 anos	22,4	87,3	12,7	100
8 a 10 anos	14,5	88,5	11,5	100
11 anos ou mais	46,1	90,3	9,7	100
Situação de Ocupação				
Ocupada	94,6	87,1	12,9	100
Desocupada	5,4	85,3	14,7	100
Renda Domiciliar Mensal percapita				
Sem renda ou até 1 s.m	49,0	82,7	17,3	100
Mais de 1 até 5 s.m	45,1	90,6	9,4	100
Mais de 5 s.m	5,9	96,1	3,9	100

Na Tabela 2 é apresentado o percentual de mulheres com 25 anos ou mais de idade, segundo as suas características comportamentais e de saúde, entre outras. A Tabela 2 fornece também o percentual de mulheres de 25 anos ou mais, segundo a realização ou não do exame Papanicolaou, para cada uma dessas características.

Na Tabela 2, observa-se que, das mulheres com 25 anos ou mais de idade, 78,1% não realizam atividade física; 68,8% nunca fumaram; e 72,3% das mulheres reportaram um estado de saúde bom ou muito bom. A maioria das mulheres (59,1%) não reportaram nenhuma das onze doenças crônicas. Com relação a *depressão*, somente 7,6% das mulheres afirmaram ter essa morbidade.

Apenas 30,8% das mulheres tem plano de saúde; 82,2% consultaram médico nos últimos 12 meses; 52,5% realizaram exame de mamografia, e 71,7% afirmaram não utilizar medicamentos de uso contínuo (MUC). Quanto as demais variáveis consideradas no estudo, verificou-se que a maioria das mulheres com 25 anos ou mais de idade, já tiveram filhos (80,8%) e possuem mãe viva (70,1%).

Entre as mulheres que praticam atividade física, apenas 7,4% não realizaram o exame preventivo para câncer de colo de útero. Esta porcentagem é praticamente o dobro (14,5%) entre as mulheres que não praticam atividade física. Com relação ao *tabagismo*, observa-se que no grupo das mulheres fumantes (14,8%), há o maior percentual de não realização do exame Papanicolaou, seguido pelo grupo das mulheres que nunca fumaram (13,2%).

Quanto a *morbidade autorreferida*, verifica-se que quanto menor o número de doenças crônicas reportadas pela mulher, maior o percentual de não realização do exame Papanicolaou. Entre as mulheres que reportaram depressão, a porcentagem das que não fizeram o exame é de 9,0%. Observa-se também que entre as mulheres que autorreferiram saúde muito boa ou boa 12,7% não realizaram o exame Papanicolaou; e entre as que autorreferiram saúde ruim ou muito ruim, o percentual de não realização do exame (15,5%) é maior.

Quanto a *mobilidade física*, há maior percentual de não realização do exame Papanicolaou, entre as mulheres que apresentam pequena dificuldade (14,2%) na realização das atividades básicas do cotidiano. Com relação ao *acesso e utilização dos serviços de saúde*, verificou-se que existe maior percentual de não realização do exame Papanicolaou entre as mulheres que não tem plano de saúde (16,3%). Entre as que não fizeram consulta médica nos 12 últimos meses, 28,5% não fizeram o exame preventivo.

Além disso, o percentual de não realização do exame é maior no grupo das mulheres que não utilizam medicamento de uso contínuo (14,4%), do que no grupo das mulheres utilizam esse tipo de medicamento (9,3%).

Com relação ao *exame de mamografia*, observou-se que entre as mulheres que fizeram a mamografia, 3,1% não realizaram o exame Papanicolaou, entretanto, entre aquelas que não fizeram a mamografia, essa porcentagem de não realização do exame Papanicolaou é muito maior (23,8%). A porcentagem de não realização do exame é praticamente o mesmo entre as mulheres que tem mãe viva (12,7%) e que não tem mãe viva (13,6%). Já a porcentagem de não realização do exame Papanicolaou entre mulheres que não tiveram filhos (22,6%) é praticamente duas vezes maior do que entre as mulheres que tiveram filhos (10,7%).

Tabela 2: Distribuição percentual de mulheres com 25 anos ou mais de idade, segundo características comportamentais, de saúde e outras no Brasil, e pelo percentual de condição de realização do exame preventivo para câncer de colo do útero (Papanicolaou).

Características	Mulheres com 25 anos ou mais	Exame Papanicolaou		Total
		Realizou o exame	Não realizou o exame	
Atividade Física				
Sim	21,9	92,6	7,4	100
Não	78,1	85,5	14,5	100
Tabagismo				
Fumante	16	85,2	14,8	100
Ex-Fumante	15,2	89,9	10,1	100
Nunca Fumou	68,8	86,8	13,2	100
Morbidade Autorreferida				
Nenhuma	59,1	85,4	14,6	100
1 doença crônica	24,3	89,2	10,8	100
2 doenças crônicas	10,2	89,4	10,6	100
3 ou mais doenças crônicas	6,4	90,6	9,4	100
Depressão				
Sim	7,6	91	9	100
Não	92,4	86,7	13,3	100

(Continuação).

Características	Mulheres com 25 anos ou mais	Exame Papanicolaou		Total
		Realizou o exame	Não realizou o exame	
Autoavaliação de Saúde				
Muito bom ou bom	72,3	87,3	12,7	100
Regular	24,3	86,7	13,3	100
Ruim ou muito ruim	3,5	84,5	15,5	100
Plano de Saúde				
Sim	30,8	94,6	5,4	100
Não	69,2	83,7	16,3	100
Consulta Médica nos Últimos 12 Meses				
Sim	82,2	90,4	9,6	100
Não	17,8	71,5	28,5	100
MUC				
Sim	28,3	90,7	9,3	100
Não	71,7	85,6	14,4	100
Limitação				
Tem dificuldade (pequena)	2	85,8	14,2	100
Não tem dificuldade	98	87,1	12,9	100
Mamografia				
Sim	52,5	96,9	3,1	100
Não	47,5	76,2	23,8	100
Mãe Viva				
Sim	70,1	87,3	12,7	100
Não	29,9	86,4	13,6	100
Teve Filhos				
Sim	80,8	89,3	10,7	100
Não	19,2	77,4	22,6	100

Na Tabela 3 são apresentados os resultados dos testes de significância geral para os parâmetros do modelo logístico binário explicativo da não realização do exame Papanicolaou, considerando todas as variáveis explicativas.

Tabela 3: Resultados dos testes de significância geral para os parâmetros do modelo logístico binário explicativo da não realização do exame Papanicolaou, considerando todas as variáveis explicativas (dezenove).

Variáveis	P-valor
Faixa Etária	<0,001
Cor / Raça	0,608
Região de Residência	0,001
Área de Localização de Domicílio	0,001
Anos de Estudo	<0,001
Situação de Ocupação	0,717
Renda Domiciliar Mensal <i>Percapita</i>	<0,001
Atividade Física	<0,001
Tabagismo	<0,001
Morbidade Autorreferida	<0,001
Depressão	0,595
Autoavaliação de Saúde	0,270
Plano de Saúde	<0,001
Consulta Médica nos Últimos 12 Meses	<0,001
MUC	0,077
Limitação Física	0,731
Mamografia	<0,001
Mãe Viva	0,656
Teve Filhos	<0,001

* *Teste F de Wald*

Com base no teste de significância geral (Tabela 3), observa-se que as variáveis *cor/raça*, *situação de ocupação*, *depressão*, *autoavaliação de saúde*, *utilização de medicamento de uso contínuo (MUC)*, *limitação na realização de atividades básicas do cotidiano* e *mãe viva*, não estão estatisticamente relacionadas com a chance da mulher não realizar o exame preventivo para o câncer do colo do útero ($p\text{-valor}>0,05$).

Tabela 4: Resultados do ajuste do modelo de regressão logística binária explicativo da não realização do exame Papanicolaou considerando todas as variáveis explicativas.

Características	Exame Papanicolaou		
	Estimativa	P-valor*	Razão de Chance
Faixa Etária			
25 a 29 anos	-1,300	<0,001	0,273
30 a 39 anos	-1,645	<0,001	0,193
40 a 49 anos	-1,437	<0,001	0,238
50 a 59 anos	-1,171	<0,001	0,310
60 a 69 anos	-0,598	<0,001	0,550
70 anos ou mais	0	-	1
Cor / Raça			
Não Branca	0,019	0,608	1,019
Branca	0	-	1
Região de Residência			
Norte	-0,271	<0,001	0,763
Nordeste	-0,036	0,595	0,964
Sudeste	-0,016	0,821	0,985
Sul	-0,056	0,455	0,945
Centro-Oeste	0	-	1
Área de Localização de Domicílio			
Urbana	-0,16	0,001	0,852
Rural	0	-	1
Anos de Estudo			
Sem instrução e menos de 1 ano	0,876	<0,001	2,401
1 a 3 anos	0,462	<0,001	1,587
4 a 7 anos	0,211	<0,001	1,235
8 a 10 anos	0,161	0,001	1,175
11 anos ou mais	0	-	1
Situação de Ocupação			
Ocupada	0,024	0,717	1,024
Desocupada	0	-	1
Renda Domiciliar Mensal per capita			
Sem renda ou até 1 s.m	0,688	<0,001	1,989
Mais de 1 até 5 s.m	0,400	0,001	1,492
Mais de 5 s.m	0	-	1

* *Teste Wald de Significância Individual.*

(Continuação).

Características	Exame Papanicolaou		
	Estimativa	P-valor*	Razão de Chance
Atividade Física			
Sim	-0,383	<0,001	0,682
Não	0	-	1
Tabagismo			
Fumante	0,052	0,204	1,053
Ex-Fumante	-0,265	<0,001	0,767
Nunca Fumou	0	-	1
Morbidade Autorreferida			
Nenhuma	0,212	0,009	1,236
1 doença crônica	0,001	0,985	1,001
2 doenças crônicas	0,022	0,796	1,022
3 ou mais doenças crônicas	0	-	1
Depressão			
Sim	-0,038	0,595	0,963
Não	0	-	1
Autoavaliação de Saúde			
Muito bom ou bom	-0,137	0,11	0,872
Regular	-0,110	0,189	0,895
Ruim ou muito ruim	0	-	1
Plano de Saúde			
Sim	-0,564	<0,001	0,569
Não	0	-	1
Consulta Médica nos Últimos 12 Meses			
Sim	-0,966	<0,001	0,381
Não	0	-	1
MUC			
Sim	-0,084	0,077	0,919
Não	0	-	1
Limitação			
Tem dificuldade (pequena)	-0,050	0,731	0,951
Não tem dificuldade	0	-	1
Mamografia			
Sim	-1,982	<0,001	0,138
Não	0	-	1
Mãe Viva			
Sim	0,018	0,656	1,018
Não	0	-	1
Teve Filhos			
Sim	-1,327	<0,001	0,265
Não	0	-	1

Os resultados encontrados a partir do teste de significância geral (Tabela 3) são confirmados pelos resultados do teste de significância individual dos parâmetros do modelo (Tabela 4). Em ambos os testes, verificou-se que não existe diferença significativa nas chances de não realização do exame Papanicolaou entre mulheres brancas e não brancas; ocupadas e não ocupadas; portadoras e não portadoras de depressão; com diferentes níveis de autoavaliação de saúde; usuárias e não usuárias de medicamento de uso contínuo; portadoras e não portadoras de limitação física; e entre mulheres com e sem mãe viva.

Após o primeiro ajuste do modelo, foram realizados novos ajustes, retirando uma a uma as variáveis que não se mostraram estatisticamente associadas (Teste de significância geral) com a chance de não realização do exame Papanicolaou, em ordem decrescente da magnitude do p-valor. Assim, excluiu-se primeiramente a variável (Limitação Física) referente ao maior p-valor, e reajustou-se o modelo sem esta variável. Em seguida, excluiu-se a próxima variável do modelo, com o maior p-valor entre as variáveis que ainda continuaram não apresentando efeito significativo; e assim sucessivamente, até a obtenção do modelo final (todas as variáveis com efeito significativo, isto é, com $p\text{-valor} \leq 0,05$). Na Tabela 5, são fornecidas as variáveis que permaneceram no modelo.

Tabela 5: Resultados dos testes de significância geral para os parâmetros do modelo logístico binário explicativo da não realização do exame Papanicolaou, considerando as doze variáveis explicativas selecionadas.

Variáveis	P-valor*
Faixa Etária	<0,001
Região de Residência	0,002
Área de Localização de Domicílio	0,001
Anos de Estudo	<0,001
Renda Domiciliar Mensal <i>per capita</i>	<0,001
Atividade Física	<0,001
Tabagismo	<0,001
Morbidade Autorreferida	<0,001
Plano de Saúde	<0,001
Consulta Médica nos Últimos 12 Meses	<0,001
Mamografia	<0,001
Teve Filhos	<0,001

**Teste F de Wald.*

A Tabela 6 fornece os resultados do ajuste do modelo logístico binário (modelo final), que incluem as estimativas pontuais dos parâmetros do modelo, as medidas de razão de chance (RC) e os p-valores do teste de Wald de significância individual.

Tabela 6: Resultados do ajuste do modelo de regressão logística binária explicativo da não realização do exame Papanicolaou, considerando todas as doze variáveis explicativas selecionadas (*modelo final*).

Características	Exame Papanicolaou		
	Estimativa	P-valor*	Razão de Chance
Faixa Etária			
25 a 29 anos	-1,263	<0,001	0,283
30 a 39 anos	-1,611	<0,001	0,200
40 a 49 anos	-1,407	<0,001	0,245
50 a 59 anos	-1,151	<0,001	0,316
60 a 69 anos	-0,588	<0,001	0,556
70 anos ou mais	0	-	1
Região de Residência			
Norte	-0,258	0,001	0,773
Nordeste	-0,03	0,661	0,971
Sudeste	-0,021	0,763	0,98
Sul	-0,065	0,386	0,937
Centro-Oeste	0	-	1
Área de Localização de Domicílio			
Urbana	-0,16	0,001	0,852
Rural	0	-	1
Anos de Estudo			
Sem instrução e menos de 1 ano	0,883	<0,001	2,418
1 a 3 anos	0,467	<0,001	1,596
4 a 7 anos	0,213	<0,001	1,237
8 a 10 anos	0,162	0,001	1,176
11 anos ou mais	0	-	1
Renda Domiciliar Mensal per capita			
Sem renda ou até 1 s.m	0,695	<0,001	2,004
Mais de 1 até 5 s.m	0,405	0,001	1,499
Mais de 5 s.m	0	-	1

* *Teste Wald de Significância Individual.*

(Continuação).

Características	Exame Papanicolaou		
	Estimativa	P-valor*	Razão de Chance
Atividade Física			
Sim	-0,388	<0,001	0,678
Não	0	-	1
Tabagismo			
Fumante	0,055	0,182	1,056
Ex-Fumante	-0,264	<0,001	0,768
Nunca Fumou	0	-	1
Morbidade Autorreferida			
Nenhuma	0,236	0,002	1,266
1 doença crônica	0,012	0,878	1,012
2 doenças crônicas	0,028	0,742	1,028
3 ou mais doenças crônicas	0	-	1
Plano de Saúde			
Sim	-0,567	<0,001	0,567
Não	0	-	1
Consulta Médica nos Últimos 12 Meses			
Sim	-0,972	<0,001	0,378
Não	0	-	1
Teve Filhos			
Sim	-1,328	<0,001	0,265
Não	0	-	1
Mamografia			
Sim	-1,985	<0,001	0,137
Não	0	-	1

* *Teste Wald de Significância Individual.*

Na Tabela 6, observou-se que a chance das mulheres não realizarem o exame Papanicolaou aumenta a partir de 30 anos de idade. Além disso, verificou-se que as mulheres de “25 a 29 anos” possuem uma chance de não realizar o exame maior que a das mulheres nas faixas etárias de “30 a 39 anos” e de “40 a 49 anos”, e menor que a das mulheres nas demais faixas etárias.

Com relação a *região de residência*, as mulheres que residem na região Centro-Oeste tem uma chance 29,4% maior de não realizarem o exame Papanicolaou do que as mulheres residentes na região Norte ($RC=1/0,773=1,294$; $p\text{-valor}=0,001$). Quanto a *área de localização de domicílio*, as mulheres que residem na área rural tem uma chance 17,4% maior de não realizarem o exame, comparativamente as mulheres na área urbana ($RC=1/0,852=1,174$; $p\text{-valor}=0,001$).

As variáveis *anos de estudo* e *renda domiciliar per capita*, apresentaram gradientes com a chance das mulheres não realizarem o exame, isto é, mulheres menos escolarizadas e com menor renda tem uma chance maior de não realizar o exame Papanicolaou.

As mulheres que não realizam atividades físicas possuem uma chance 47,5% maior de não realizar o exame preventivo para câncer de colo de útero, quando comparada às mulheres que praticam alguma atividade física ($RC=1/0,678=1,475$; $p\text{-valor}<0,001$). Quanto a variável *tabagismo*, verificou-se que as mulheres não fumantes têm uma chance 30,2% maior de não realizar o exame do que as mulheres ex-fumantes ($RC=1/0,768=1,302$; $p\text{-valor}<0,001$). Já as mulheres fumantes tem uma chance 5,6% maior de não realizar o exame Papanicolaou do que as mulheres não fumantes, mas esta diferença não foi estatisticamente significativa ($RC=1,056$; $p\text{-valor}=0,182$).

As mulheres sem nenhuma doença crônica (considerando as 10 doenças crônicas utilizadas na construção desse indicador de morbilidade autorreferida) possuem uma chance 26,6% maior de não realizarem o exame Papanicolaou do que as mulheres que reportaram três ou mais doenças crônicas ($RC=1,266$; $p\text{-valor}=0,002$).

Verificou-se que as variáveis *plano de saúde* e *consulta médica* estão fortemente associadas com a chance das mulheres não realizarem o exame Papanicolaou (Tabela 6). As mulheres que não possuem plano de saúde possuem uma chance 76,4% maior de não realizarem o exame, comparativamente as mulheres que possuem plano de saúde ($RC=1/0,567=1,764$; $p\text{-valor}<0,001$). Já as mulheres que não realizaram consulta médica nos últimos doze meses possuem uma chance 2,6 vezes maior de não realizarem o exame Papanicolaou, em comparação as mulheres que realizaram consulta médica neste período ($RC=1/0,378=2,64$; $p\text{-valor}<0,001$).

Para a variável *teve filho* observa-se que as mulheres que não tiveram filho tem uma chance 3,8 vezes maior de não realizar o exame Papanicolaou ($RC=1/0,265=3,774$; $p\text{-valor}<0,001$), do que as mulheres que tiveram filho. Com relação a variável *mamografia*, observou-se que as mulheres que não realizaram a mamografia possuem uma chance 7,3 vezes maior de não realizarem o exame Papanicolaou, comparativamente aquelas mulheres que realizaram o exame de mamografia ($RC=1/0,137=7,30$; $p\text{-valor}<0,001$). O fato da mulher não ter tido filho e não ter feito a mamografia estão fortemente associadas com a chance de não realização do exame Papanicolaou.

No que se refere à qualidade do ajuste do modelo final, observou-se que 88,2% das mulheres foram classificadas corretamente pelo modelo. Além disso, as mulheres que de fato realizaram o exame Papanicolaou, observou-se que 98,4% foram classificadas corretamente pelo modelo; enquanto que, das mulheres que não realizaram o exame, 19,0% foram classificadas corretamente. Segundo os *pseudos-R²* de *McFadden* e *Nagelkerke* o ganho de informação obtido com este modelo (modelo final), em comparação ao modelo nulo, foi de 23,6% e 30,9%, respectivamente.

A Figura 2 apresenta a curva ROC. A área sob a curva ROC é 0,834, indicando que a capacidade discriminatória do modelo é boa.

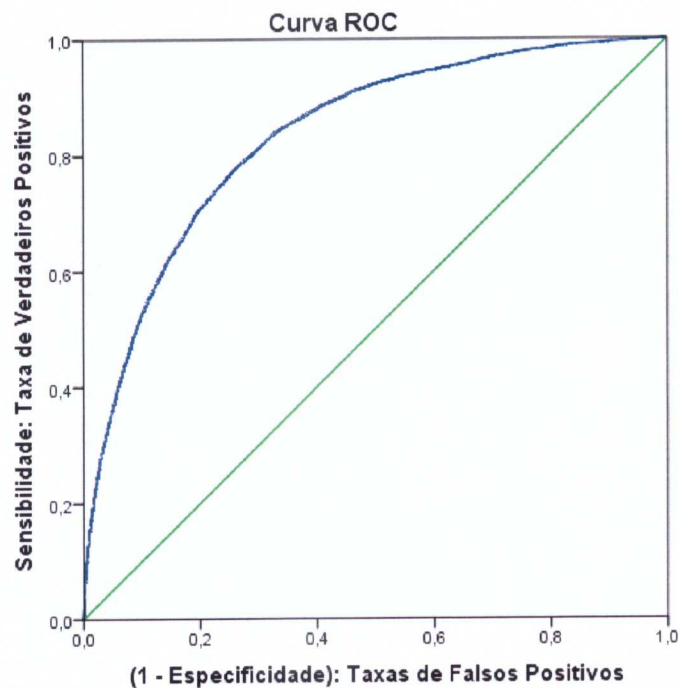


Figura 2: Área sob curva ROC.

4.2. Discussão

O câncer de colo de útero é uma neoplasia que apresenta um bom prognóstico se diagnosticado e tratado precocemente; desta forma a abordagem preconizada para o controle populacional consiste na realização do rastreamento através do exame preventivo para câncer do colo do útero, também conhecido como exame de Papanicolaou, ou ainda como exame colpocitológico. O teste citopatológico convencional (Papanicolaou) é a principal estratégia de programas de rastreamento do câncer do colo do útero no mundo. No Brasil, a estratégia recomendada pelo Ministério da Saúde é que este exame seja realizado em mulheres de 25 a 64 anos.

A neoplasia citada no presente estudo é a terceira mais incidente na população feminina brasileira. Impulsionado pelo Programa Viva Mulher, criado em 1996, o controle do câncer do colo do útero foi reafirmado como prioridade no *plano de fortalecimento da rede de prevenção, diagnóstico e tratamento do câncer*, lançado pela presidente da República, em 2011. Para a efetividade do programa de controle do câncer do colo do útero, faz-se necessário garantir a organização, a integralidade e a qualidade dos serviços e ações da linha de cuidado, bem como o tratamento e o seguimento das pacientes.

O presente estudo objetivou verificar a associação entre algumas características socioeconômicas, demográficas, comportamentais e de saúde das mulheres, de 25 anos ou mais de idade, e a chance de não realização do exame Papanicolaou.

Segundo o estudo de OLIVEIRA *et al.* (2006), considerando o grupo de mulheres na faixa-etária de 25 a 49 anos, quanto maior for a idade maior será a chance de não realização o exame Papanicolaou. Neste estudo verificou-se também essa tendência, mas a partir de 30 anos de idade, já que as mulheres de 25 a 29 anos apresentaram uma chance de não realização do exame maior que as mulheres nas faixas etárias de 30 a 39 anos e 40 a 49 anos. Com relação a escolaridade, ALBUQUERQUE *et al.* (2009) e OLIVEIRA *et al.* (2006) também concluíram que mulheres mais escolarizadas possuem menor chance de não realização do exame Papanicolaou. A associação encontrada entre a renda e a chance de não realização do exame preventivo para câncer colo do útero, também foi verificada por BORGES *et al.* (2012) e OLIVEIRA *et al.* (2006), indicando que mulheres com menores rendimentos tem uma chance maior de não realizar o exame preventivo.

Assim como ALBUQUERQUE *et al.* (2009) e OLIVEIRA *et al.* (2006), também foi observado que o fato da mulher ter consultado o médico e o fato de já ter tido filho estão associados a uma menor chance de não realização do exame para câncer do colo de útero. No presente estudo, um outro fator associado com a não realização do exame Papanicolaou é a posse de plano de saúde, isto é, observou-se que mulheres com acesso a plano de saúde tem uma chance maior de realizar o exame preventivo, comparativamente as mulheres sem acesso a plano de saúde. Este resultado também foi encontrado em diferentes estudos, como os de MORAES *et al.* (2011) e GASPERIN *et al.* (2011).

Foi verificado que mulheres que não realizaram mamografia tem uma chance maior de não realizar o exame preventivo de câncer de colo de útero. Como neste estudo, BORGES *et al.* (2012), também concluíram que mulheres com alguma morbidade tem chance maior de realizar o exame preventivo, em comparação com aquelas que não tem nenhuma morbidade.

Os estudos de NOVAES *et al.* (2006) e SOUZA *et al.* (2008), mostraram que mulheres da zona rural tem uma chance maior de não realizar o exame preventivo comparado as mulheres da área urbana. Neste presente estudo também foi verificado essa associação, além de encontrar uma menor chance de não realização do exame Papanicolaou em mulheres residentes na região Sudeste.

Embora no presente estudo não exista diferença estatisticamente significativa entre mulheres fumantes e as que nunca fumaram, o resultado encontrado mostrou que mulheres fumantes realizam menos o exame Papanicolaou do que as mulheres que nunca fumaram. Este sentido da associação é condizente com o estudo de OLIVEIRA *et al.* (2006). Com relação a mulheres ex-fumantes, neste trabalho também foi verificado que as mulheres ex-fumantes realizam mais o exame Papanicolaou do que as mulheres que nunca fumaram.

Mulheres que praticam atividades físicas realizam mais o exame preventivo do que aquelas que não se exercitam. Essa associação também foi encontrada por BORGES *et al.* (2012), mas essa associação só foi significativa na análise bruta.

5. Conclusões e Considerações Finais

A partir dos resultados deste trabalho pode-se concluir que mulheres mais velhas, que moram na área rural, menos escolarizadas e com menor renda domiciliar per capita são as que menos fizeram o exame preventivo de câncer de colo de útero (Papanicolaou). Conclui-se também que mulheres que praticam alguma atividade física, são portadoras de maior número de doenças crônicas, tem acesso a plano de saúde, fizeram pelo menos uma consulta médica no último ano, tiveram filhos e fizeram o exame de mamografia são as que mais fizeram o exame Papanicolaou.

Outros estudos foram realizados visando estabelecer a associação entre um conjunto de fatores e a não realização do exame preventivo para câncer do colo de útero, usando diferentes indicadores, como tabagismo, morbidade, plano de saúde, atividade física, se teve filhos, etc. Na maioria desses estudos foram utilizados dados de algumas localidades e o indicador de morbidade é dicotomizado (nenhuma doença crônica *versus* pelo menos uma doença crônica). Neste sentido, este trabalho buscou levantar informações de abrangência nacional, investigando a associação da não realização do exame preventivo com características demográficas, socioeconômicas, comportamentais e de saúde da mulher no Brasil.

Baseados nos resultados deste estudo conclui-se que existe a necessidade de sensibilizar determinados grupos de mulheres, especialmente aquelas menos escolarizadas e com baixo rendimento econômico, sobre a importância do exame Papanicolaou; isto poderia ser feito, por exemplo, por meio de educação em saúde. Além disso, é necessário universalizar o acesso aos serviços de saúde nos ambientes urbano e rural e em diferentes regiões brasileiras. Cabe por último ressaltar a necessidade de investir na capacitação de profissionais de atenção primária, e agilizar o diagnóstico tanto na rede pública quanto na rede privada.

Referências Bibliográficas

- ALBUQUERQUE, K.M; FRIAS, P.G; ANDRADE, C.L.T; AQUINO, E.M.L; MENEZES, G; SZWARCOWALD, C.L. **Cobertura do teste de Papanicolaou e fatores associados à não-realização: um olhar sobre o Programa de Prevenção do Câncer do Colo do Útero em Pernambuco, Brasil.** Cad. Saúde Pública, Rio de Janeiro, 25 Sup 2:S301-S309, 2009.
- ALDRIGHI, J.M; ALDRIGHI, A.P.S; PETTA, C.A. **Contracepção hormonal oral, HPV e risco de câncer cérvico-uterino.** Rev. Assoc. Med. Bras. [online]. 2002, 48(2): pp.96-96. ISSN 0104-4230.
- AMORIM, V.M.S.L; BARROS, M.B.A; CÉSAR, C.L.G; CARANDINA, L; GOLDBAUM, M. **Fatores associados à não realização do exame de Papanicolaou: um estudo de base populacional no Município de Campinas, São Paulo, Brasil.** Cad. Saúde Pública, Rio de Janeiro, 22(11): 2329-2338, nov, 2006
- ANDRADE, M.G; ALMEIDA, M.M.G; ARAÚJO, T.M.; SANTOS, K.O.B. **Fatores associados a não adesão ao Papanicolaou entre mulheres atendidas pela Estratégia Saúde da Família em Feira de Santana, Bahia, 2010.** Epidemiol. Serv. Saúde, Brasília, 23(1): 111-120, jan-mar 2014.
- ASPAROUHOV, T. **Sampling weights in latent variable modeling.** Structural Equation Modeling, 12(3): 411-434, 2005.
- BINDER, D.A. **On the variance of asymptotically normal estimators from complex survey.** International Statistical Review, 51, 279-292; 1983.
- BORGES, M.F.S.O; DOTTO, L.M.G; KOIFMAN, R.J; CUNHA, M.A; MUNIZ, P.T. **Prevalência do exame preventivo de câncer do colo do útero em Rio Branco, Acre, Brasil, e fatores associados à não-realização do exame.** Cad. Saúde Pública, Rio de Janeiro, 28(6):1156-1166, jun, 2012.
- BRASIL. Ministério da Saúde. Instituto Nacional de Câncer. **Prevenção do câncer do colo do útero: normas e recomendações do INCA.** Revista Brasileira de Cancerologia. Rio de Janeiro, 49(4): 205, out./dez. 2003.
- BRASIL. Ministério da Saúde. Secretaria de Ciência e Tecnologia e Insumos Estratégicos. Departamento de Ciência e Tecnologia. **Integração de informações dos registros de câncer brasileiros.** Revista de Saúde Pública, 41(5): 865-868, out. 2007.
- BRASIL. Portaria GM/MS nº 399, de 22 de dezembro de 2006. Dispõe sobre normas para publicação no Diário Oficial e no Diário da Justiça. Diário Oficial [da] República Federativa do Brasil, Brasília, DF, 16 dez. 2002. Disponível em: <http://dtr2001.saude.gov.br/sas/PORTARIAS/Port2006/GM/GM-399.htm>
- CARVALHO, F.B; RODRIGUES, D.A; SANTOS, N.R. **Fatores relevantes a não realização do exame Papanicolaou em acadêmicas de enfermagem da UNIGRAN.** Interbio 5(2), 2011 - ISSN 1981-3775.

DUAVY, L.M; BATISTA, F.L; JORGE, M.S.B; SANTOS, J.B.F. **A percepção da mulher sobre o exame preventivo do câncer cérvico-uterino: estudo de caso**; Ciência & Saúde Coletiva, 12(3):733-742, 2007.

FERREIRA, J.M.O; BRIRRO, F.B. **Análise Espacial da Incidência de Câncer do Colo do Útero no Município do Rio de Janeiro no Período de 1995 a 1998**. (Graduação em Estatística)— Escola Nacional de Ciências Estatísticas, Rio de Janeiro, 2006.

GASPERIN, S; BOING, A.F; KUPEK, E. **Cobertura e fatores associados à realização do exame de detecção do câncer de colo de útero em área urbana no Sul do Brasil: estudo de base populacional**. Cad. Saúde Pública, Rio de Janeiro, 27(7): 1312-1322, jul, 2011.

HOSMER, D.W; LEMESHOW, S. **Applied Logistic Regression** (2nd Edition). Jonh Wiley & Sons, New York. (2000).

IARC-International Agency for Research on Cancer[homepage on the internet]. **Confirms efficacy of cervix cancer screening for women 25-65 in reducing mortality**. Press release nº 151. Lyon; 2004 [cited 2004 May 3]. Available from: http://www.iarc.fr/ENG/Press_Releases/Summary.pdf

IBGE. Pesquisa nacional por amostra de domicílios 2008: notas metodológicas - pesquisa básica, pesquisa especial de tabagismo e pesquisas suplementares de saúde e acesso à internet e posse de telefone móvel celular para uso pessoal, Rio de Janeiro: IBGE, 2009. Disponível em: <http://www.ibge.gov.br>.

INCA. ABC DO CÂNCER – **Abordagens Básicas para o Controle de Câncer**, 2ª edição. 2012. Rio de Janeiro, 2012.

Instituto Nacional de Câncer José Alencar Gomes da Silva. **Estimativa 2012**: incidência de câncer no Brasil [Internet]. Rio de Janeiro: INCA; 2011.118p. Disponível em: <http://www.inca.gov.br/estimativa/2012/estimativa20122111.pdf>

Instituto Nacional de Câncer José Alencar Gomes da Silva. **Estimativa 2014**: incidência de câncer no Brasil [Internet]. Rio de Janeiro: INCA, 2014. Disponível em: <http://www.inca.gov.br/estimativa/2014/estimativa-24042014.pdf>

JORGE, R.J.B; SAMPAIO, L.R.L; DIÓGENES, M.A.R, MENDONÇA, F.A.C; SAMPAIO, L.L. **Fatores associados a não realização periódica do exame Papanicolaou**. Rev Rene, Fortaleza, 2011 jul/set; 12(3):606-12.

MATOS, D.O. **O Papel da enfermagem na busca pela adesão das mulheres ao exame de papanicolaou**. Universidade Federal de Minas Gerais, 2011.

MARTINS, L.F.L; THULER, L.C.S; VALENTE, J.G. **Cobertura do exame de Papanicolaou no Brasil e seus fatores determinantes: uma revisão sistemática da literatura**. Rev. Bras. Ginecol. Obstet. 27(8): 485-92, 2005.

MARTINS, L.F.L; VALENTE, J.G; THULER, L.C.S. **Fatores associados à não**

realização do exame de Papanicolaou: estudo transversal de base populacional em duas capitais brasileiras. Revista Brasileira de Cancerologia 52(2): 197, 2006.

MORAES, J.R.; GUIMARÃES, P.V.; PAULA, F.L.; FERREIRA, M.L.P; GUIMARÃES, R.M.; LUIZ, R.R. **Relação entre plano de saúde e a realização do exame Papanicolaou: uma aplicação de escore de propensão usando um inquérito amostral complexo.** Rev. Bras. Epidemiol., 14(4): 589-97, 2011.

MORAES, J.R.; MOREIRA, J.P.L.; LUIZ, R.R. **Associação entre o estado de saúde autorreferido de adultos e a área de localização do domicílio: uma análise de regressão logística ordinal usando a PNAD 2008.** Ciência & Saúde Coletiva [online], 16(9), 3769-3780, 2011.

MORAES, J.R. de, MOREIRA, J.P.L., LUIZ, R.R. Efeito do plano amostral em modelo logístico ordinal: uma análise do estado de saúde autorreferido de adultos no Brasil usando a Pesquisa Nacional por Amostra de Domicílios de 2008. Cadernos de Saúde Pública, 28(5), 913-924, 2012.

NOVAES, H.; BRAGA, P.; SCHOUT, D. Fatores associados à realização de exames preventivos para câncer nas mulheres brasileiras, PNAD 2003. Ciência & Saúde Coletiva, 11(4), 2006.

OLIVEIRA, A.F; CUNHA, C.L.F; VIÉGAS, I.F; FIGUEIREDO, I.S; BRITO, L.M.O; CHEIN, M.B.C. **Estudo sobre a adesão ao exame citopatológico de papanicolaou em um grupo de mulheres.** Rev Pesq Saúde, 11(1): 32-37, jan-abr, 2010.

OLIVEIRA, M.M; SILVA, A.A.; BRITO, L; COIMBRA, L. Cobertura e fatores associados à realização do exame preventivo de Papanicolaou em São Luis, Maranhão. Revista Brasileira de Epidemiologia, 9(3), 325-334, 2006.

OLIVEIRA, F.C.R; KRETZMANN, C.K. **Fatores determinantes da realização de exames preventivos de saúde da mulher no Brasil, usando a PNAD 2008.** 9ª mostra acadêmica UNIMEP, 2011.

PESSOA, D.G.C; SILVA, P. L. **Análise estatística de dados da PNAD: incorporando a estrutura do plano amostral.** São Paulo: ABE, 1998.

PESSOA, D.G.C; SILVA, P.L.N; LILÁ, M.F. **Análise estatística de dados da PNAD: incorporando o plano amostral.** Ciência & Saúde Coletiva, 7(4): 659-670, 2002.

PRADO, P.R; KOIFMAN, R.J; SANTANA, A.L.M; SILVA, I.F. **Rastreamento do CCU em Rio Branco.** Revista Brasileira de Cancerologia 58(3): 471-479, 2012.

SOUZA, L.M; FIORAVANTE, E. **Fatores associados à realização dos exames preventivos de câncer de mama e de colo uterino, pelas mulheres brasileiras.** Trabalho apresentado no XVI Encontro Nacional de Estudos Populacionais, realizado em Caxambu- MG –Brasil, de 29 de setembro a 03 de outubro de 2008.

SZWARCWALD, C.L; DAMACENA, G.N. **Amostras complexas em inquéritos populacionais: planejamento e implicações na análise estatística dos dados.** Revista Brasileira de Epidemiologia, 11(1): 38-45, 2008.

TAVARES, C.M.A.; PRADO, M.L. Pesquisando a Prevenção do Câncer Ginecológico em Santa Catarina. Contexto Enfermagem, Florianópolis, 15(4): 578-586, 2006.

THULER, L.C.S. **Câncer do Colo do Útero no Estado de Mato Grosso do Sul: Detecção Precoce, Incidência e Mortalidade.** Revista Brasileira de Cancerologia 58(3): 399, 2012.

WHO (World Health Organization). **Cytopatological Screening in the Control of Cervical Cancer. Geneva: WHO; 1988.**

WHO (World Health Organization). **Manual on the prevention and control of common cancers. Geneva: WHO; 1998.** (Westerns Pacific Series, 20).

ZEFERINO, L.C. **O desafio de reduzir a mortalidade por câncer do colo do útero.** Rev. Bras Ginecol Obstet. 30(5): 213-215, 2008.