



Universidade de Brasília - UnB
Instituto de Ciências Exatas - IE
Departamento de Estatística - EST

Trabalho de Conclusão de Curso

Análise de Variáveis Categóricas de Resposta Múltipla

Leonardo Maciel Levenhagen Pinto
16/0033802

Orientador: Prof. Dr. George Freitas von Borries

Brasília

2019

Leonardo Maciel Levenhagen Pinto
16/0033802

Trabalho de Conclusão de Curso
Análise de Variáveis Categóricas de Resposta Múltipla

Trabalho de Conclusão de Curso como um dos requisitos para a obtenção do título de Bacharel em Estatística da Universidade de Brasília.

Orientador: Prof. Dr. George Freitas von Borries

Brasília

2019

Leonardo Maciel Levenhagen Pinto

Análise de Variáveis Categóricas de Resposta Múltipla / Leonardo Maciel
Levenhagen Pinto. – Brasília, 2019.

116p.

Orientador: Prof. Dr. George Freitas von Borries

Relatório Final – Universidade de Brasília

Instituto de Ciências Exatas

Departamento de Estatística

Trabalho de Conclusão de Curso de Graduação, 2019.

1. Introdução. 2. Metodologia. 3. Aplicações. 4. Conclusão.

Leonardo Maciel Levenhagen Pinto
16/0033802

Trabalho de Conclusão de Curso
Análise de Variáveis Categóricas de Resposta Múltipla

Trabalho de Conclusão de Curso como um dos requisitos para a obtenção do título de Bacharel em Estatística da Universidade de Brasília.

Trabalho aprovado. Brasília, 12 de dezembro de 2019:

Prof. Dr. George Freitas von Borries
Orientador

Prof. Dr. Eduardo Freitas da Silva
Convidado 1

Prof. Ma. Maria Teresa Leão Costa
Convidado 2

Brasília
2019

Este trabalho é dedicado a mim

Agradecimentos

Agradeço a todos os meus professores da universidade pelas experiências repassadas e conhecimentos ensinados. Em especial, agradeço às professoras Ana Maria Nogales Vasconcelos, Cátia Regina Gonçalves, Claudete Ruas, Maria Teresa Leão Costa e aos professores Alan Ricardo da Silva, Caio Cesar de Medeiros Costa, Diego Marques Ferreira e Eduardo Freitas da Silva. Agradeço também aos meus amigos dos quais emprestaram-me seus ouvidos quando eu precisava desabafar.

*“Só há progresso
quando se derruba o retrocesso.”
(Eduardo D. Ramos C. Matos)*

Resumo

Questionários com perguntas cujas respostas podem ser marcações de nenhum a todos os itens são recorrentes. As variáveis associadas à essas perguntas são chamadas de variáveis categóricas de resposta múltipla (MRCV) e testar a independência entre essas variáveis é comum para pesquisadores. No entanto, como um mesmo respondente pode marcar mais de um item para uma mesma pergunta, há dependência entre as respostas e testes tradicionais envolvendo a estatística Qui-Quadrado de Pearson são inadequadas. Neste trabalho, a literatura sobre MRCV é revisada e aplicada no questionário de Percepção sobre Computação do projeto Meninas da Computação conduzido no Departamento de Ciência da Computação da UnB. O ajuste de Bonferroni, a estatística Qui-Quadrado *bootstrap* aproximada por uma distribuição Qui-Quadrado adequada, o produto dos valores-p de *bootstrap*, o menor valor-p de *bootstrap*, o ajuste de segunda ordem de Rao-Scott e o modelo log-linear generalizado são sugestões dos autores. A análise dos dados advindos do questionário é feita no *software* R usando o pacote MRCV. Para o período de 2011 a 2014, dependência estatística entre a área de interesse das respondentes com o local de uso do computador, a área de interesse com as atividades realizadas no computador e o local de uso e atividades realizadas foram encontradas. Para 2019, dependência significativa apenas entre local de uso e atividades realizadas foi encontrada.

Palavras-chave: independência marginal. modelo log-linear. resposta múltipla. variável categórica.

Abstract

Survey questions whose answers could be more than one marking are recurring. Variables associated with those questions are called multiple-response categorival variables (MRCV) and testing independence between these variables is common among researchers. However, one participant could mark more than one item for a single question, causing dependence between answers which in turn invalidates Pearson chi-square traditional tests. In this paper, literature about MRCV is reviewed and applied to the survey *Percepção sobre Computação* of *Meninas da Computação* project which is conducted in University of Brasilia's department of computer science. Bonferroni adjustment, bootstrap chi-square approximated by an adequate distribution, bootstrap product of p-values, bootstrap minimum p-value, second-order adjusted chi-square and log-linear model are suggestions from the authors. The analysis of the survey data is done in software R using the MRCV package. From 2011 to 2014, statistic dependence between participants' area of interest with places where they use the computer, participants' area of interest with activities done in the computer and places where they use the computer with activities done in the computer have been found. As for 2019, statistic dependence only between places where they use the computer with activities done in the computer have been found.

Keywords: marginal independence. log-linear model. multiple response. categorical variable.

Lista de ilustrações

Figura 1 – Exemplo de variáveis categóricas de resposta múltipla (MRCVs) no questionário Percepção sobre Computação do projeto Meninas na Computação.	30
Figura 2 – Formas gerais de tabelas de contingência para contagens, probabilidades e frequências esperadas.	44
Figura 3 – Gráficos de mosaico para as variáveis Área de Interesse vs. Local de Uso de Computador e Área de Interesse vs. Atividade Realizada no Computador (2011 - 2014).	53
Figura 4 – Gráfico de mosaico para as variáveis Área de Interesse e Local de Uso de Computador (2019).	59

Lista de tabelas

Tabela 1	– Exemplo de uma tabela de contingência marginal para uma variável categórica de única resposta (SRCV) e uma variável categórica de resposta múltipla (MRCV): 6 cultivares de morango e 16 atributos do tipo <i>Check All That Apply</i>	34
Tabela 2	– Exemplo de uma tabela de contingência marginal para uma variável categórica de única resposta (SRCV) e uma variável categórica de resposta múltipla (MRCV) condicionada a uma outra SRCV: Tabela marginal para dados UTI	34
Tabela 3	– Exemplo de uma tabela de contingência marginal para duas variáveis categóricas de resposta múltipla (MRCVs): Estimativas da proporção com pesos de pesquisa para os dados de NHANES (2008).	35
Tabela 4	– Tabela de resposta ao item das estimativas da proporção com pesos de pesquisa para os dados de NHANES (2008).	36
Tabela 5	– Forma geral de uma tabela de resposta ao item	47
Tabela 6	– Respostas para o questionário Percepção sobre Computação do projeto Meninas na Computação	51
Tabela 7	– Tabela de contingência marginal para as variáveis Área de Interesse e Local de Uso de Computador (2011 - 2014).	53
Tabela 8	– Tabela de contingência marginal para as variáveis Área de Interesse e Atividade Realizada no Computador (2011 - 2014).	54
Tabela 9	– Saída R com as estatísticas Qui-Quadrado de Pearson não ajustadas e valores-p ajustados por Bonferroni para as variáveis Área de Interesse vs. Local de Uso do Computador (2011-2014).	55
Tabela 10	– Saída R com as estatísticas Qui-Quadrado de Pearson não ajustadas e valores-p ajustados por Bonferroni para as variáveis Área de Interesse vs. Atividade Realizada no Computador (2011-2014).	55
Tabela 11	– Valores-p para os testes MMI e SPMI (2011-2014).	55
Tabela 12	– Valores-p por ajuste de Bonferroni para cada estatística $X_{S(i,j)}^2$ para o teste SPMI das variáveis Local de Uso de Computador e Atividade Realizada no Computador (2011 - 2014)	56
Tabela 13	– Valores-p por ajuste de Bonferroni para cada estatística $X_{S(i,j)}^2$ para o teste SPMI das variáveis Local de Uso de Computador e Atividade Realizada no Computador para interessadas em Exatas (2011 - 2014)	56

Tabela 14 – Valores-p por ajuste de Bonferroni para cada estatística $X_{S(i,j)}^2$ para o teste SPMI das variáveis Local de Uso de Computador e Atividade Realizada no Computador para interessadas em Biológicas e Saúde (2011 - 2014)	57
Tabela 15 – Valores-p por ajuste de Bonferroni para cada estatística $X_{S(i,j)}^2$ para o teste SPMI das variáveis Local de Uso de Computador e Atividade Realizada no Computador para interessadas em Humanas (2011 - 2014)	57
Tabela 16 – Tabela de contingência marginal para as variáveis Área de Interesse e Local de Uso de Computador (2019).	59
Tabela 17 – Tabela de contingência marginal para as variáveis Área de Interesse e Atividade Realizada no Computador (2019)	60
Tabela 18 – Saída R com as estatísticas Qui-Quadrado de Pearson não ajustadas e valores-p ajustados por Bonferroni para as variáveis Área de Interesse vs. Local de Uso do Computador (2019).	61
Tabela 19 – Saída R com as estatísticas Qui-Quadrado de Pearson não ajustadas e valores-p ajustados por Bonferroni para as variáveis Área de Interesse vs. Atividade Realizada no Computador (2019).	61
Tabela 20 – Valores-p para os testes MMI e SPMI (2019).	61
Tabela 21 – Valores-p por ajuste de Bonferroni para cada estatística $X_{S(i,j)}^2$ para o teste SPMI das variáveis Local de Uso de Computador e Atividade Realizada no Computador (2019)	62
Tabela 22 – Valores-p por ajuste de Bonferroni para cada estatística $X_{S(i,j)}^2$ para o teste SPMI das variáveis Local de Uso de Computador e Atividade Realizada no Computador para interessadas em Exatas (2011 - 2019) .	62
Tabela 23 – Valores-p por ajuste de Bonferroni para cada estatística $X_{S(i,j)}^2$ para o teste SPMI das variáveis Local de Uso de Computador e Atividade Realizada no Computador para interessadas em Biológicas e Saúde (2019)	63
Tabela 24 – Valores-p por ajuste de Bonferroni para cada estatística $X_{S(i,j)}^2$ para o teste SPMI das variáveis Local de Uso de Computador e Atividade Realizada no Computador para interessadas em Humanas (2019)	63
Tabela 25 – Estatísticas de teste de máxima verossimilhança <i>bootstrap</i> para comparação de modelos log-lineares marginais generalizados para as variáveis Local de Uso do Computador e Atividades Realizadas no Computador.	65
Tabela 26 – Modelos ajustados, seus graus de liberdade, sua estatística de razão de verossimilhança e seu AIC comparado ao modelo saturado.	66
Tabela 27 – Tabela de resposta ao item para as variáveis Área de Interesse e Local de Uso de Computador (2011 - 2014).	73
Tabela 28 – Tabela de resposta ao item para as variáveis Área de Interesse e Local de Uso de Computador (2019).	73

Tabela 29 – Tabela de resposta ao item para as variáveis Área de Interesse e Atividade Realizada no Computador (2011 - 2014).	74
Tabela 30 – Tabela de resposta ao item para as variáveis Área de Interesse e Atividade Realizada no Computador (2019).	74
Tabela 31 – Tabela de contingência marginal para as variáveis Local de Uso de Computador e Atividade Realizada no Computador (2011 - 2014).	75
Tabela 32 – Tabela de resposta ao item para as variáveis Local de Uso de Computador e Atividade Realizada no Computador (2011 - 2014).	76
Tabela 33 – Tabela de contingência marginal para as variáveis Local de Uso de Computador e Atividade Realizada no Computador para interessadas em Exatas (2011 - 2014).	77
Tabela 34 – Tabela de contingência marginal para as variáveis Local de Uso de Computador e Atividade Realizada no Computador para interessadas em Biológicas e Saúde (2011 - 2014).	78
Tabela 35 – Tabela de contingência marginal para as variáveis Local de Uso de Computador e Atividade Realizada no Computador para interessadas em Humanas (2011 - 2014).	79
Tabela 36 – Tabela de contingência marginal para as variáveis Locais de Uso de Computador e Atividade Realizada no Computador (2019).	80
Tabela 37 – Tabela de resposta ao item para as variáveis Locais de Uso de Computador e Atividade Realizada no Computador (2019).	81
Tabela 38 – Tabela de contingência marginal para as variáveis Local de Uso de Computador e Atividade Realizada no Computador para a área de interesse em Exatas (2019).	82
Tabela 39 – Tabela de contingência marginal para as variáveis Local de Uso de Computador e Atividade Realizada no Computador para a área de interesse em Biológicas e Saúde (2019).	83
Tabela 40 – Tabela de contingência marginal para as variáveis Local de Uso de Computador e Atividade Realizada no Computador para a área de interesse em Humanas (2019).	84

Lista de abreviaturas e siglas

SRCV	<i>Single-response categorical variable</i>
MRCV	<i>Multiple-response categorical variable</i>
CATA	<i>Choose all that apply</i>
MMI	<i>Multiple Marginal Independence</i>
SPMI	<i>Simultaneous Pairwise Marginal Independence</i>
UnB	Universidade de Brasília
NANHES	<i>National Health and Nutrition Examination Survey</i>

Lista de símbolos

π_{ij}	Probabilidade conjunta de respostas positivas a X_i e Y_j
π_{i+}	Probabilidade marginal de respostas positivas para $X_i = 1$
π_{+j}	Probabilidade marginal de respostas positivas para $Y_j = 1$
$\hat{\pi}_{ij}$	Estimador de máxima verossimilhança para π_{ij}
$\hat{\pi}_{i+}$	Estimador de máxima verossimilhança para π_{i+}
$\hat{\pi}_{+j}$	Estimador de máxima verossimilhança para π_{+j}
m_{ij}	Número de respostas positivas para $X_i = 1$ e $Y_j = 1$
m_{i+}	Número de respostas positivas para $X_i = 1$
m_{j+}	Número de respostas positivas para $Y_j = 1$
m_{++}	Tamanho amostral
μ_{ij}	Frequência esperada de respostas positivas para $X_i = 1$ e $Y_j = 1$
μ_{i+}	Frequência esperada de respostas positivas para $X_i = 1$
μ_{j+}	Frequência esperada de respostas positivas para $Y_j = 1$
$X_{S(i,j)}^2$	Estatística de Pearson tradicional para uma tabela (i, j) de contingência 2×2
X_M^2	Estatística do Teste para MMI
X_S^2	Soma de estatísticas de Pearson tradicionais
α	Erro do tipo I
α_E	Erro experimental
α_C	Erro comparativo
p_{ij}	Valor-p para cada $X_{S(i,j)}^2$
p^*	Valor-p ajustado de Westfall e Young (1993)
F	Função de distribuição de probabilidade
\hat{F}	Estimador de F

\widehat{F}_n	Distribuição empírica de F
$\gamma(F)$	Funcional
ϑ	Característica populacional de interesse
$\widehat{\vartheta}$	Estimador de ϑ
$\widetilde{\vartheta}$	Estimador corrigido de ϑ
$b(\widehat{\vartheta})$	Viés de $\widehat{\vartheta}$
$\widehat{b}(\widehat{\vartheta})$	Estimador de $b(\widehat{\vartheta})$
$\widehat{\vartheta}_I^*$	Percentil 100 ($\alpha/2$) da distribuição bootstrap de $\widehat{\vartheta}$
$\widehat{\vartheta}_S^*$	Percentil 100 ($1 - \alpha/2$) da distribuição bootstrap de $\widehat{\vartheta}$
$X_{M(b)}^2$	Estatística do Teste <i>bootstrap</i> para MMI
$X_{S(b)}^2$	Estatística do Teste <i>bootstrap</i> para SPMI
X_{RS2}^2	X_S^2 com ajuste de segunda ordem de Rao-Scott
λ	Efeito geral
λ_i^X	Efeito marginal da variável X
λ_j^Y	Efeito marginal da variável Y
λ_k^Z	Efeito marginal da variável Z
λ_{ij}^{XY}	Efeito de interação das variáveis X e Y
λ_{ik}^{XZ}	Efeito de interação das variáveis X e Z
λ_{jk}^{YZ}	Efeito de interação das variáveis Y e Z
λ_{ijk}^{XYZ}	Efeito de interação das variáveis X , Y e Z
e_{ijk}	Resíduo de Pearson
h_{ijk}	Ponto de alavanca
\mathfrak{P}	População
\mathfrak{s}	Amostra de \mathfrak{P}
p_w	Pesos de pesquisa
$m_{uv(ij)}$	Contagem marginal populacional para $(x_i = u, y_j = v)$

$\mu_{uv(ij)}$	Estimador de $m_{uv(ij)}$
$\mu_{uv(ij)}^{(0)}$	Total populacional predito sob H_0
$\mu_{uv(ij)}^{(A)}$	Total populacional predito sob H_A
$\pi_{uv(ij)}$	Proporção marginal populacional para $(x_i = u, y_j = v)$
$p_{uv(ij)}$	Estimador de $\pi_{uv(ij)}$
$p_{uv(ij)}^{(0)}$	Proporção populacional predito sob H_0
$p_{uv(ij)}^{(A)}$	Proporção populacional predito sob H_A
$r_{uv(ij)}$	Resíduo de Pearson padronizado

Sumário

1	PROBLEMA DE RESPOSTAS MÚLTIPLAS	29
1.1	Introdução	29
1.2	Revisão Bibliográfica	30
1.3	Objetivos	32
2	METODOLOGIA	33
2.1	Tabelas de Contingência	33
2.2	Teste de hipótese para independência marginal	37
2.2.1	Ajuste de Bonferroni	39
2.2.2	Aproximação por <i>bootstrap</i>	40
2.2.3	Correções de Rao-Scott	42
2.3	Modelo log-linear marginal generalizado	43
3	APLICAÇÃO	51
3.1	Respostas de 2011 a 2014	52
3.2	Respostas de 2019	58
3.3	Ajuste do modelo	64
4	CONCLUSÃO	67
	REFERÊNCIAS	69
	APÊNDICES	71
	APÊNDICE A – TABELAS DE RESPOSTA AO ITEM	73
	APÊNDICE B – CÓDIGO DO PROGRAMA R	85

1 Problema de Respostas Múltiplas

1.1 Introdução

O uso de variáveis categóricas é recorrente em questionários formulados por diversos campos de estudo como agricultura, agronomia, ciência de alimentos, educação, psicologia, e sociologia, entre outras áreas. Essas variáveis categóricas apresentadas em questionários geralmente são respondidas em questões com marcação de apenas um entre c itens possíveis. No entanto, a análise de variáveis categóricas que permitem a escolha de múltiplos itens como resposta, chamadas *multiple-response categorical variables* (MRCVs) ou variáveis categóricas de resposta múltipla, também é uma conjuntura habitual em problemas envolvendo a aplicação de questionários. Dados com a presença desse tipo de variável são geralmente denominados *Choose all that apply data* ou *Check all that apply data*, resumidos pela expressão *CATA data* ou dados CATA.

Como os participantes podem marcar mais de um item, é plausível a possibilidade de dependência entre as respostas. Quando surge esse tipo de problema em que há pelo menos uma variável categórica de resposta múltipla, Loughin e Scherer (1998), Agresti e Liu (1999), Thomas e Decady (2004) e Bilder e Loughin (2004) recomendam não recorrer aos métodos de análise de variáveis categóricas de resposta única (SRCVs - *single-response categorical variables*) como os testes Qui-Quadrado tradicionais de independência de Pearson. Identifica-se por meio desses trabalhos citados um interesse por uma independência específica, a independência marginal, e a inadequabilidade de modelos amostrais usuais para as contagens das respostas advindas de variáveis categóricas de resposta múltipla resultante das contagens não serem independentes, não possuem distribuição de probabilidade de Poisson e também não serem multinomiais.

A partir das referências de Bilder e Loughin (2004) e Bilder e Koziol (2014) e com base nas recomendações de não aplicar métodos de análise de independência para variáveis categóricas de resposta singular em variáveis categóricas de resposta múltipla, desperta-se interesse em testes de independência para respostas múltiplas, titulados por Agresti e Liu (1999) como *Simultaneous Pairwise Marginal Independence* (SPMI) ou testes para independência marginal em pares simultâneos, e modelos para dados *Check all that apply*.

Uma aplicação será realizada com dados do questionário de Percepção sobre Computação, do projeto Meninas na Computação (de Holanda, M.T., et al., 2011 - atual) conduzido no Departamento de Ciência da Computação da Universidade de Brasília. A Figura 1 apresenta a parte do questionário usado nesse problema com variáveis categóricas de resposta múltipla geradoras dos dados do tipo *Check all that apply*.

Figura 1 – Exemplo de variáveis categóricas de resposta múltipla (MRCVs) no questionário Percepção sobre Computação do projeto Meninas na Computação.

PERCEPÇÃO SOBRE COMPUTAÇÃO

1. Marque todos os locais em que você usa o computador:

<input type="radio"/> em casa <input type="radio"/> em casa de parentes <input type="radio"/> em casa de amigos <input type="radio"/> na escola	<input type="radio"/> no trabalho <input type="radio"/> em lan-house <input type="radio"/> biblioteca <input type="radio"/> centros de inclusão digital
--	--

2. Marque todas as atividades que você realiza ou realizou no computador:

<input type="radio"/> edição de texto (word, outros) <input type="radio"/> edição de imagem (outros) <input type="radio"/> planilha (excel, outros) <input type="radio"/> banco de dados (access, outros) <input type="radio"/> acesso a internet (pesquisas sobre conteúdos, notícias, outros) <input type="radio"/> redes sociais (facebook, orkut, etc)	<input type="radio"/> e-mail <input type="radio"/> jogos <input type="radio"/> desenvolvimento de páginas <input type="radio"/> programação <input type="radio"/> outros.
---	---

Fonte: Holanda et al. (2011 - atual).

1.2 Revisão Bibliográfica

As terminologias *multiple-response categorical variables* e *CATA data* e suas traduções livres empregadas nesse trabalho não são as primeiras tentativas de nomear bancos de dados e variáveis onde participantes podem escolher mais de um item como resposta para cada pergunta de um determinado questionário, tampouco são tentativas recentes. Coombs (1964) utiliza a nomenclatura *pick any data* para se referir aos dados provenientes de questionários do tipo selecione todos os itens aplicáveis. Sob sugestão de Coombs (1962) e Coombs e Smith (1973), Levine (1979) e Umesh (1995) denominam esses dados de diferentes formas: *pick any data*, *pick k out of n data* e *pick any out of n data*. Vagamente, Nettleton (1999) menciona as expressões *mark all that apply questions* e *multiple-response data*. Decady e Thomas (2000) mencionam tanto a expressão *multiple-response data* quanto *pick any data* em seu trabalho. Com uma pequena modificação na notação, Bilder e Loughin (2002a, 2002b) utilizam *pick any out of c data* e *pick any out of c variables* embora em trabalhos posteriores Bilder e Loughin (2003, 2004, 2007, 2009) e Bilder e Koziol (2014) escolhem *multiple-response categorical variables* e *multiple-response categorical data* como terminologia padrão. Meyners e Castura (2016) associam aos dados à nomenclatura *CATA data* como alternativa às opções anteriores.

Apesar de tantas formas para fazer referência a um mesmo banco de dados que permitem diversos itens como resposta, todas as expressões são muito símeis. As expressões *pick any data*, *pick any out of n data*, *multiple-response data*, *pick any out of c data*, *multiple-response categorical data* e *CATA data* podem ser traduzidas livremente como dados do tipo escolha quaisquer itens, dados do tipo escolha quaisquer itens dentre n itens disponíveis, dados de resposta múltipla, dados do tipo escolha quaisquer itens dentre c itens disponíveis,

dados de resposta múltipla, dados categóricos de resposta múltipla e, finalmente, dados do tipo escolha todas as opções aplicáveis, respectivamente. Equivalentemente, *mark all that apply questions*, *pick any out of c variables* e *multiple-response categorical variables* também são traduzíveis livremente como perguntas do tipo marque todas as opções aplicáveis, variáveis do tipo escolha quaisquer itens dentre c itens disponíveis e variáveis categóricas de resposta múltipla.

O grande arsenal de nomenclaturas referente a esse tipo de banco de dados pode ser problemático considerando uma possível confusão nas pesquisas por artigos e livros da mesma temática e igual confundimento ao interpretar erroneamente as diferentes terminologias como referências a diferentes bancos de dados. Com o intuito de explicar claramente a análise proposta nesse trabalho, apenas as nomenclaturas dados CATA e variáveis categóricas de resposta múltipla serão utilizadas por serem mais recentes.

A magnitude da análise de variáveis categóricas de resposta múltipla data do final do século XX. Umesh (1995) utiliza uma estatística de Pearson modificada e sua aproximação por uma distribuição Qui-Quadrado. No entanto, essa abordagem é inválida dada a dependência entre as respostas múltiplas para um indivíduo qualquer. Loughin e Scherer (1998) analisam grau de escolaridade e fontes primárias de informação veterinária, ou seja, uma variável categórica de resposta única (SRCV) e uma variável de resposta múltipla (MRCV) via *bootstrap* da estatística de Pearson modificada de Umesh (1995) para testar independência em tabelas de contingência. Agresti e Liu (1999) sugerem alternativas para os mesmos dados por ajuste de máxima verossimilhança ou quasi-verossimilhança, um ajuste de Bonferroni e a soma de estatísticas de Pearson tradicionais além de um modelo logito marginal para testar *Multiple Marginal Independence* ou independência marginal múltipla. Decady e Thomas (1999) testam homogeneidade múltipla marginal por meio da estatística de Pearson modificada com correção de primeira ordem de Rao-Scott. Nettleton (1999) revisa as metodologias sugeridas pelos autores mencionados, apontando invariância da estatística do teste como um empecilho.

A partir do século XXI, Bilder e Loughin (2001) apontam novamente a invariância da estatística Umesh-Loughin-Scherer como um impasse na análise correta de dados CATA além de problemas com forte dependência pareada nas respostas dos indivíduos. Bilder e Loughin (2002a) testam *Conditional Multiple Marginal Independence*, ou independência marginal múltipla condicional, entre uma SRCV e uma MRCV condicionadas a uma outra SRCV por aproximação da estatística de Cochran modificada por uma distribuição Qui-Quadrado, *bootstrap* dessa mesma estatística, *bootstrap* do mínimo e do produto dos valores-p e ajuste de Bonferroni. De forma parecida, Bilder e Loughin (2002b, 2004) realizam um teste para *Simultaneous Pairwise Marginal Independence* (SPMI), ou independência marginal em pares simultaneamente, entre duas MRCVs com as mesmas técnicas exceto pelo fato de usarem estatísticas de Pearson modificadas e o uso da correção de segunda

ordem de Rao-Scott para tal estatística modificada. Bilder e Loughin (2003) ajustam um modelo log-linear generalizado estimado por máxima verossimilhança, um modelo log-linear marginal generalizado e um modelo logito marginal estimado por *generalized estimating equation*, ou equações de estimação generalizadas para testar SPMI. Bilder e Loughin (2007, 2009) sugerem um modelo para duas ou mais MRCVs, calculam as razões de chances observadas e preditas pelo modelo e seus intervalos de confiança bem como diferentes tipos de associação permitidas pelo modelo ajustado.

1.3 Objetivos

Este trabalho faz uma breve revisão bibliográfica sobre a metodologia acerca de dados do tipo *check all that apply* e análise de um problema de pesquisa prático utilizando o pacote MRCV do *software* R. A análise aborda testes de independência marginal múltipla (MMI) e independência marginal em pares simultâneos (SPMI) para os dados do questionário Percepção sobre Computação. Este trabalho está organizado na seguinte forma. O Capítulo 2 aborda a metodologia empregada neste trabalho, incluindo tabelas de contingência marginal e de resposta ao item, testes de hipóteses para testar independência marginal múltipla (MMI) e independência marginal em pares simultâneos (SPMI) que são extensões da estatística Qui-Quadrado de Pearson e modelo log-linear generalizado. O Capítulo 3 descreve a aplicação dessa metodologia no questionário Percepção sobre Computação usando o pacote MRCV do *software* R. O Capítulo 4 expõe as conclusões e observações pertinentes.

2 Metodologia

2.1 Tabelas de Contingência

Uma tabela de contingência é um tipo de tabela de dupla entrada na qual se observa a distribuição das frequências das variáveis. Essa tabela é uma opção exploratória de dados fundamental para um entendimento geral do problema de pesquisa e ajuda a verificar relações entre as variáveis. Segundo a literatura atual existem três tipos básicos de dados envolvendo variáveis categóricas de resposta múltipla (MRCVs) para os quais tabelas de contingência podem ser construídas. O primeiro tipo de banco de dados envolve uma variável categórica de resposta singular (SRCV) e uma MRCV. Já o segundo tipo de conjunto de dados aborda uma SRCV e uma MRCV condicionadas a uma outra SRCV. Por fim, existem dados cujas variáveis são ambas do tipo MRCV.

A Tabela 1 é um exemplo de uma tabela de contingência marginal para uma SRCV e uma MRCV. Nesta tabela, cultivar de morango é SRCV e atributos é MRCV em um estudo de análise sensorial acerca de morangos e suas características e os totais representam o número total de marcações para cada cultivar. O estudo foi feito com consumidores de um supermercado no Uruguai onde eles selecionavam os atributos associadas a cada cultivar a fim de descrevê-lo com base nos próprios sentidos. Como um consumidor marca mais de um atributo para a caracterização de um mesmo morango e algumas opções são mutualmente exclusivas para descrição dos cultivares de morangos como pequeno e grande, existe dependência entre o consumidor e os atributos.

A Tabela 2 é um exemplo de uma tabela de contingência marginal para uma SRCV e uma MRCV condicionada a uma outra SRCV. Os dados relacionados a essa tabela são de histórico de mulheres com infecção do trato urinário (sim ou não) e contraceptivos usados, condicionado a sua idade ($<$ ou \geq 24 anos) e tanto os totais de respostas marcadas como o número de mulheres respondentes para cada grupo de idade estabelecido são apresentados. Note que diferentes contraceptivos podem ser utilizados pela mesma mulher.

A Tabela 3 é um exemplo de uma tabela de contingência marginal pelo fato de mostrar apenas as respostas positivas para as categoriais associadas. As proporções estimadas apresentadas se referem às MRCVs uso de tabaco durante toda a vida (\geq 100 cigarros, \geq 20 cachimbos, \geq 20 charutos, \geq rapés e \geq tabacos de mascar) e problemas respiratórios (tosse frequente, retirada de fleuma frequente, chiado ou silvo no peito e tosse seca à noite) no ano de 2008 tratados pelo NHANES (*National Health and Nutrition Examination Survey*), conduzido pela primeira vez em 1971 e aplicado anualmente nos Estados Unidos desde 1999. Diferentemente das outras duas tabelas, um mesmo indivíduo

pode ter usado mais de um tipo de tabaco durante toda a vida e ter mais de um problema respiratório.

Tabela 1 – Exemplo de uma tabela de contingência marginal para uma variável categórica de única resposta (SRCV) e uma variável categórica de resposta múltipla (MRCV): 6 cultivares de morango e 16 atributos do tipo *Check All That Apply*.

Cultivar	Atributos																Total
	Doce	Amargo	Gosto de morango	Odor de morango	Saboroso	Sem gosto	Cor vermelha	Formato irregular	Formato regular	Pequeno	Grande	Firme	Duro	Macio	Suculento	Seco	
Yurí	29	16	19	19	25	43	50	25	36	21	52	55	37	14	52	23	516
Yvahé	39	14	21	21	38	40	40	30	35	39	28	49	13	32	74	13	526
Guenoa	35	16	22	25	45	32	64	17	44	18	54	44	5	49	76	13	559
Festival	37	30	21	32	50	23	54	25	31	14	47	70	22	22	62	19	559
L20.1	52	10	24	31	59	22	62	13	43	11	71	53	10	39	63	23	586
K31.5	14	38	18	16	21	41	45	11	63	16	48	57	21	18	62	24	513
Total	206	124	125	144	238	201	315	121	252	119	300	328	108	174	389	115	

Fonte: Meyners e Castura (2016). "Total" se refere ao número total de itens marcados.

Tabela 2 – Exemplo de uma tabela de contingência marginal para uma variável categórica de única resposta (SRCV) e uma variável categórica de resposta múltipla (MRCV) condicionada a uma outra SRCV: Tabela marginal para dados UTI

	UTI	Contraceptivo					Total de respostas	Total de mulheres
		Oral	Pres.	Pres. lubrificado	Espermicida	Diafragma		
Idade ≥ 24	Não	18 (0,75)	9 (0,38)	8 (0,33)	7 (0,29)	0 (0,00)	42	24
	Sim	8 (0,57)	9 (0,64)	2 (0,14)	3 (0,21)	2 (0,14)	24	14
	Total	26 (0,68)	18 (0,47)	10 (0,26)	10 (0,26)	2 (0,05)		38
Idade < 24	Não	55 (0,65)	41 (0,48)	37 (0,44)	27 (0,32)	0 (0,00)	160	85
	Sim	75 (0,65)	68 (0,59)	33 (0,28)	22 (0,19)	5 (0,04)	203	116
	Total	130 (0,65)	109 (0,54)	70 (0,35)	49 (0,24)	5 (0,02)		201

Fonte: Bilder e Loughin (2002); valores em parênteses são as proporções da seleção de um contraceptivo; Preservativo é abreviado como Pres.

Tabela 3 – Exemplo de uma tabela de contingência marginal para duas variáveis categóricas de resposta múltipla (MRCVs): Estimativas da proporção com pesos de pesquisa para os dados de NHANES (2008).

Uso de tabaco durante toda a vida	Problemas respiratórios no ano de 2008			
	Tosse na maioria dos dias durante um período de 3 meses	Retirada de fleuma na maioria dos dias durante um período de 3 meses	Chiado ou silvo no peito	Tosse seca à noite
Cigarro ≥ 100 vezes	0,0517	0,0523	0,0962	0,0248
Cachimbo ≥ 20 vezes	0,0127	0,0134	0,0204	0,0059
Charuto ≥ 20 vezes	0,0168	0,0174	0,0288	0,0092
Rapé ≥ 20 vezes	0,0055	0,0061	0,0129	0,0039
Tabaco de mascar ≥ 20 vezes	0,0064	0,0095	0,0137	0,0036

Fonte: Bilder e Loughin (2009).

Pode-se representar também esses dados por uma grande tabela contendo tabelas de contingência 2×2 , contabilizando assim as respostas positivas e negativas para cada item. Essa tabela maior é chamada de *tabela de resposta ao item* e a partir dela inferências são construídas. Seja X_i a resposta binária (1 = positiva, 0 = negativa) para o item da linha $i = 1, \dots, I$ e seja Y_j a resposta binária para o item da coluna $j = 1, \dots, J$. Na Tabela 3, $X_1 = 1$ indica um indivíduo que fumou pelo menos 100 cigarros em sua vida toda, $Y_1 = 1$ indica um indivíduo com tosse na maioria dos dias durante um período de 3 meses. Considerando a combinação de níveis (X_1, Y_1) das variáveis X e Y , temos que $(X_1 = 1, Y_1 = 1)$ indica um indivíduo que fumou pelo menos 100 cigarros em sua vida toda e tosse na maioria dos dias durante o período estipulado, $(X_1 = 1, Y_1 = 0)$ indica um indivíduo que fumou pelo menos 100 cigarros em sua vida toda, mas não tossiu na maioria dos dias, $(X_1 = 0, Y_1 = 1)$ indica um indivíduo que não fumou pelo menos 100 cigarros em sua vida toda, mas tossiu na maioria dos dias e $(X_1 = 0, Y_1 = 0)$ indica um indivíduo que nem fumou pelo menos 100 cigarros em sua vida toda nem tossiu na maioria dos dias. Repetindo o processo considerando cada combinação de itens como feito acima, o resultado é a construção de IJ tabelas de contingência 2×2 . A Tabela 4 é a tabela de resposta ao item e reproduz as $IJ = 20$ tabelas de contingência possíveis para os dados de NHANES. Fazendo um paralelo dessa representação mais completa da Tabela 4 com a Tabela 3, as células da Tabela 3 são proporções relacionadas às contagens para a resposta combinada $(X_i = 1, Y_i = 1)$ nas $IJ = 20$ tabelas possíveis enquanto as células da Tabela 4 apresentam todas as combinações de níveis (X_i, Y_j) para $i = 1, \dots, I$ e $j = 1, \dots, J$.

Tabela 4 – Tabela de resposta ao item das estimativas da proporção com pesos de pesquisa para os dados de NHANES (2008).

	Problemas respiratórios no ano de 2008								
	Tosse na maioria dos dias durante um período de 3 meses		Retirada de fleuma na maioria dos dias durante um período de 3 meses		Chiado ou silvo no peito		Tosse seca à noite		
	1	0	1	0	1	0	1	0	
Uso de tabaco durante toda a vida									
Cigarro \geq 100 vezes									
1	0,0517	0,4427	0,0523	0,4420	0,0962	0,3982	0,0248	0,4695	
0	0,0196	0,4860	0,0175	0,4882	0,0500	0,4556	0,0125	0,4931	
Cachimbo \geq 20 vezes									
1	0,0127	0,0796	0,0134	0,0789	0,0204	0,0719	0,0059	0,0864	
0	0,0586	0,8491	0,0565	0,8513	0,1258	0,7819	0,0315	0,8763	
Charuto \geq 20 vezes									
1	0,0168	0,1278	0,0174	0,1273	0,0288	0,1159	0,0092	0,1355	
0	0,0545	0,8009	0,0524	0,8029	0,1174	0,7379	0,0282	0,8272	
Rapé \geq 20 vezes									
1	0,0055	0,0481	0,0061	0,0475	0,0129	0,0407	0,0039	0,0497	
0	0,0658	0,8806	0,0637	0,8827	0,1333	0,8131	0,0334	0,9130	
Tabaco de mascar \geq 20 vezes									
1	0,0064	0,0511	0,0095	0,0480	0,0137	0,0438	0,0036	0,0539	
0	0,0650	0,8775	0,0603	0,8822	0,1325	0,8100	0,0338	0,9087	

Fonte: Bilder e Louglin (2009)

É de suma importância notar as diferenças da representação encontrada na Tabela 3, de uma tabela de contingência usual. Não há representação das proporções estimadas relativas às contagens para as respostas $(X_i = 1, Y_j = 0)$, $(X_i = 0, Y_j = 1)$ e $(X_i = 0, Y_j = 0)$, impedindo uma boa visualização do quadro geral dos dados. Outra diferença já comentada é a possibilidade de responder mais de um item para as perguntas levantadas. Por conta disso, faz-se necessária a introdução de uma análise adequada levando em consideração essas diferenças como teste de hipóteses adequados e ajuste de modelos.

2.2 Teste de hipótese para independência marginal

Na análise de variáveis categóricas de resposta múltipla, o foco principal é determinar se a probabilidade de uma resposta positiva para um item altera com base na resposta de outras perguntas. A resposta a esta questão pode ser obtida através de testes de independência marginal em pares simultâneos. Para a Tabela 3, por exemplo, gostaríamos de saber se a probabilidade de uso de alguma forma de tabaco seria dependente do surgimento de problemas respiratórios específicos. Generalizando, seria cada tabela de contingência 2×2 da tabela de resposta ao item, independente uma da outra? Pelo fato de essas tabelas de contingência 2×2 resumirem os dados marginalmente e o interesse ser o teste de independência entre elas, a terminologia dada para esse teste por Agresti e Liu (1999) é o teste para independência marginal em pares simultâneos (SPMI).

Testes para independência marginal múltipla (MMI) entre uma variável categórica de única resposta (SRCV) e uma variável categórica de resposta múltipla (MRCV) são investigados por Umesh (1995) e Loughin e Scherer (1998), refinados por Agresti e Liu (1999) e Bilder et al. (2000). Testar para independência marginal em pares condicional (CPMI) entre uma SRCV e uma MRCV condicionado por uma outra variável do tipo SRCV é visto em Bilder e Loughin (2002). Testes para independência marginal em pares simultâneos (SPMI) são examinados por Thomas e Decady (2000) e Bilder e Loughin (2004), não obstante limitados a esses autores. O interesse deste trabalho reside nos testes para MMI e SPMI.

Para testar MMI, considere $\mathbf{Y} = (Y_1, \dots, Y_J)$ como a variável categórica de resposta múltipla onde $Y_j = \{0, 1\}$ para os itens $j = 1, \dots, J$. Como Y_j é uma resposta binária, podemos ter 2^J vetores de resposta ao item \mathbf{Y} diferentes. Para cada item $i = 1, \dots, I$ da variável categórica de resposta singular, defina $\pi_{j|i}$ como a probabilidade marginal de uma resposta positiva ($Y_j = 1$) e $\tau_{k|i}$ como a probabilidade de observar o vetor de resposta ao item $k = 1, \dots, 2^J$. Além disso, para cada item $i = 1, \dots, I$ da SRCV, a probabilidade marginal $\pi_{j|i}$ pode ser calculada pela soma das probabilidades de observar o k -ésimo vetor de resposta ao item pela equação

$$\pi_{j|i} = \sum_{\{k: Y_j=1\}} \tau_{k|i}$$

e as hipóteses do teste de independência para MMI são

$$\begin{cases} H_0 : \pi_{j|1} = \dots = \pi_{j|I} \text{ para } j = 1, \dots, J \\ H_1 : \text{Pelo menos uma das igualdades é falsa.} \end{cases}$$

Para testar independência para independência marginal em pares simultâneos (SPMI), defina $\pi_{ij} = P(X_i = 1, Y_i = 1)$ como a probabilidade conjunta pareada para $i = 1, \dots, I$ e $j = 1, \dots, J$. Defina também $\pi_{i+} = P(X_i = 1)$ e $\pi_{+j} = P(Y_j = 1)$ as probabilidades marginais das respostas positivas aos itens correspondentes, respectivamente. As hipóteses do teste de SPMI são

$$\begin{cases} H_0 : \pi_{ij} = \pi_{i+}\pi_{+j} & \text{para } i = 1, \dots, I \text{ e } j = 1, \dots, J \\ H_1 : \pi_{ij} \neq \pi_{i+}\pi_{+j} & \text{em pelo menos um par } (i, j). \end{cases}$$

Defina as estimativas de máxima verossimilhança da probabilidade conjunta pareada como $\hat{\pi}_{ij}$, da probabilidade marginal da resposta X_i como $\hat{\pi}_{i+}$ e da probabilidade marginal da resposta Y_j como $\hat{\pi}_{+j}$ para $i = 1, \dots, I$ e $j = 1, \dots, J$. Essas estimativas são dadas por $\hat{\pi}_{ij} = m_{ij}/m_{++}$, $\hat{\pi}_{i+} = m_{i+}/m_{++}$ e $\hat{\pi}_{+j} = m_{+j}/m_{++}$, onde m_{ij} é o número de respostas positivas para $(X_i = 1, Y_j = 1)$, m_{i+} o número de respostas positivas para $X_i = 1$, m_{+j} é o número de respostas positivas para $Y_j = 1$ e m_{++} é o tamanho amostral.

Por sugestão de Bilder et al. (2000), seja X_M^2 a estatística do teste para testar independência marginal múltipla (MMI) dada por

$$X_M^2 = \sum_{j=1}^J \sum_{i=1}^I m_{i+} \frac{\left(\frac{m_{ij}}{m_{i+}} - \frac{m_{+j}}{m_{++}} \right)^2}{\frac{m_{+j}}{m_{++}} \left(1 - \frac{m_{+j}}{m_{++}} \right)}.$$

A estatística do teste tem distribuição $\chi_{J(I-1)}^2$ sob MMI. Rejeita-se MMI se X_M^2 é maior do que o quantil $1 - \alpha$ da distribuição.

Seja $X_{S(i,j)}^2$ a estatística de Pearson tradicional para o teste de independência em uma das (i, j) tabelas de contingência 2×2 da tabela de resposta ao item. Bilder e Loughin (2004) sugerem como estatística do teste

$$X_S^2 = \sum_{i=1}^I \sum_{j=1}^J X_{S(i,j)}^2 = m_{++} \sum_{i=1}^I \sum_{j=1}^J \frac{(\hat{\pi}_{ij} - \hat{\pi}_{i+}\hat{\pi}_{+j})^2}{\hat{\pi}_{i+}\hat{\pi}_{+j} (1 - \hat{\pi}_{i+}) (1 - \hat{\pi}_{+j})}.$$

A princípio, o uso de X_M^2 seria usada para testar independência marginal múltipla (MMI), porém Bilder et al. (2000) explicam que X_M^2 depende de assintóticas e as contagens deveriam ser grandes o suficiente para isso. Além disso, as respostas apresentam dependência em muitas situações. Bilder et al. (2000) provam que a distribuição assintótica de X_M^2 sob MMI é uma combinação linear de $J(I-1)$ variáveis aleatórias χ_1^2 .

Em uma primeira instância, a soma de todas as estatísticas de Pearson tradicionais, X_S^2 , das tabelas de contingência 2×2 seria usada para testar a independência marginal em pares simultaneamente (SPMI) com aproximação de sua distribuição de probabilidade por χ_{IJ}^2 . No entanto, Bilder e Loughin (2001, 2004) explicam que $X_{S(i,j)}^2$ são provavelmente estatísticas não independentes porque cada um dos participantes tem sua resposta contabilizada em cada um das IJ tabelas de contingência 2×2 , podendo ser positiva ou negativa. Pelo exemplo, a relação entre usar um tipo específico de tabaco e gerar um determinado problema respiratório é factível. Na presença de respostas de um indivíduo para uma variável categórica de resposta múltipla associadas significativamente, ignorar o fato de as estatísticas não serem independentes resulta em um teste cuja rejeição da hipótese nula H_0 é demasiada frequente quando SPMI é verdadeira. Alguns procedimentos para testar MMI e SPMI adequadamente são o ajuste de Bonferroni, aproximação por *bootstrap* e correções de Rao-Scott.

2.2.1 Ajuste de Bonferroni

De forma geral, a compreensão do erro associado à comparações múltiplas é tortuosa. Esclarecendo essa situação, seja $X \in \mathbb{N}^0$ o número de erros do tipo I (α) cometidos em múltiplos testes. Assumindo independência dos testes, a probabilidade de cometer uma quantidade x de erros do tipo I é

$$P(X = x) = \binom{n}{x} \alpha^x (1 - \alpha)^{n-x},$$

a probabilidade de não cometer erros do tipo I é

$$P(X = 0) = \binom{n}{0} \alpha^0 (1 - \alpha)^{n-0} = (1 - \alpha)^n,$$

e a probabilidade de cometer pelo menos um erro do tipo I é

$$P(X \geq 1) = 1 - P(X = 0) = 1 - (1 - \alpha)^n.$$

Denomina-se a probabilidade $P(X \geq 1) = \alpha_E$ de *erro experimental* ou erro para grupo de testes e $P(X = 1) = \alpha_C$ será o *erro comparativo* ou erro para o teste individual, a probabilidade de cometer apenas um erro. Assim,

$$\alpha_E = 1 - (1 - \alpha_C)^n \text{ e } \alpha_C = 1 - (1 - \alpha_E)^{1/n}.$$

O ajuste de Bonferroni (1936) é uma correção para múltiplas comparações entre estatísticas de teste não independentes testadas simultaneamente. Sua ideia é a redução do erro do tipo I para um conjunto de hipóteses.

Agresti e Liu (1999) sugerem a aplicação do ajuste de Bonferroni nas estatísticas tradicionais de Pearson para cada uma das J sub-tabelas marginais $I \times 2$ para testar

independência marginal múltipla (MMI), rejeitando-a se pelo menos uma das estatísticas é maior do que o quantil $1 - \alpha/J$ de uma distribuição χ_{I-1}^2 .

Uma forma de contornar a invalidez da combinação das estatísticas individuais das tabelas de contingência 2×2 em uma soma é calcular o valor-p para cada tabela e combiná-las em um único teste para independência marginal em pares simultâneos (SPMI). Isso pode ser feito utilizando o ajuste de Bonferroni em testes individuais para cada tabela 2×2 . Seja α_E o nível de significância para o grupo de testes individuais e p_{ij} o valor-p para cada estatística $X_{S(i,j)}^2$ de uma tabela (i, j) para $i = 1, \dots, I$ e $j = 1, \dots, J$. Para cada estatística $X_{S(i,j)}^2$, calcula-se seu valor-p usando uma aproximação da distribuição χ_1^2 ou usando um método exato. Após calculados, rejeita-se SPMI caso $p_{ij} < \alpha_E/(IJ)$.

Uma alternativa para o teste SPMI é usar o valor-p ajustado de Westfall e Young (1993). Defina $p^* = IJ \times \min_{ij}(p_{ij})$ o valor-p ajustado de Westfall e Young (1993), onde $p^* = 1$ se o produto $IJ \times \min_{ij}(p_{ij}) > 1$. Rejeita-se SPMI se $p^* < \alpha_E$.

Mitchell e Cabin (2000) e Shinichi Nakagawa (2004) criticam o aumento no erro do tipo II ao utilizar o ajuste de Bonferroni. Em seus trabalhos, Mitchell e Cabin apontam interpretações das descobertas resultantes de um teste com o uso do ajuste de Bonferroni dependentes do número de outros testes realizados enquanto Nakagawa supõe um aumento na tendência dos pesquisadores em esconderem os resultados não esperados, encobrirem variáveis cujo resultado não fora o desejado. No problema levantado nesse trabalho, há pouca rejeição dos testes de SPMI baseados em ajustes de Bonferroni quando SPMI é verdadeira caso o número de tabelas 2×2 seja pequeno.

2.2.2 Aproximação por *bootstrap*

Seja uma população com função de distribuição de probabilidade F . Defina uma característica ou estatística populacional de interesse $\vartheta = \gamma(F)$, onde ϑ é um valor e γ é um funcional capaz de mapear funções de distribuição para \mathbb{R} . $\vartheta = \gamma(F)$ pode ser a média, a mediana, a moda, a variância, o desvio padrão, etc. Seja agora $X = \{x_1, \dots, x_n\}$ uma amostra aleatória de tamanho n advinda de uma população com função de distribuição F . Defina \hat{F} o estimador dessa função F . Bickel e Doksum (1977) definem a estimativa da característica de interesse como a aplicação do mesmo funcional γ na distribuição estimada \hat{F} como $\hat{\vartheta} = \gamma(\hat{F})$. Se quiséssemos, por exemplo, a média de uma distribuição F univariada como a característica de interesse $\vartheta = \gamma(F)$, poderíamos escolher como o estimador de ϑ :

$$\hat{\vartheta} = \gamma(\hat{F}) = \frac{1}{n} \sum_{i=1}^n x_i = \bar{x}.$$

Algumas perguntas acerca do quão bom é esse estimador $\hat{\vartheta}$ naturalmente são levantadas. A estimativa fornecida está próxima de ϑ ? O estimador é não viesado? Sua variabilidade é baixa? Qual a distribuição de $\hat{\vartheta}$?

A obtenção de uma característica populacional ϑ de interesse pode ser laboriosa, seja por aproximações via teoria estatística não trivial, seja por sua complexidade matemática. Uma solução inicial é retirar inúmeras amostras da população e simular uma aproximação da função de distribuição de probabilidade de $\hat{\vartheta}$. Contudo, há um custo associado à amostragem e apenas uma amostra é fornecida. Nesses casos, o *bootstrap* é uma boa ferramenta para estimar tal característica.

Usando a notação anterior, o método de *bootstrap* usa a amostra $X = \{x_1, \dots, x_n\}$ de F para estimar ϑ pelo seu estimador $\hat{\vartheta} = \gamma(\hat{F})$, onde \hat{F} é o estimador da distribuição de probabilidade de F . Posteriormente, a abordagem de simulação com \hat{F} no lugar de F para encontrar uma aproximação da distribuição de $\hat{\vartheta}$.

Uma forma de obter \hat{F} para estimar F é uma abordagem paramétrica, assumindo conhecida uma distribuição F com parâmetros a serem estimados por procedimentos padrões. Outra forma para a obtenção de \hat{F} é uma abordagem não-paramétrica, uma primeira consideração para \hat{F} é a função de distribuição empírica acumulada de F , função cuja atribuição de massa de probabilidade é $1/n$ para cada elemento da amostra X . Todavia frisa-se o fato de essa escolha não ser obrigatória apesar de intuitiva e conveniente em diversas situações. A distribuição empírica de F é dada por

$$\hat{F}_n = \frac{1}{n} \sum_{i=1}^n I(X_i \leq x),$$

onde $I(X_i \leq x)$ é a função indicadora

$$I(X_i \leq x) = \begin{cases} 1 & \text{se } X_i \leq x \\ 0 & \text{caso contrário} \end{cases}$$

Pode-se também estimar e corrigir o viés usando a abordagem *bootstrap*. Defina

$$b(\hat{\vartheta}) = E(\hat{\vartheta}) - \vartheta$$

como o viés da estatística de interesse. O estimador do viés dessa estatística considerando a distribuição empírica para \hat{F} é definido por

$$\hat{b}(\hat{\vartheta}) = E(\hat{\vartheta}) - \hat{\vartheta} \approx \frac{1}{B} \sum_{b=1}^B \hat{\vartheta}_b^* - \hat{\vartheta}$$

com correção do viés estimado em

$$\tilde{\vartheta} = \hat{\vartheta} - \hat{b}(\hat{\vartheta}) = \hat{\vartheta} - \left(\frac{1}{B} \sum_{b=1}^B \hat{\vartheta}_b^* - \hat{\vartheta} \right) = 2\hat{\vartheta} - \frac{1}{B} \sum_{b=1}^B \hat{\vartheta}_b^*.$$

Um cuidado ao tomar essa correção $\tilde{\vartheta}$ para o viés de $\hat{\vartheta}$ é verificar se a variância de $\tilde{\vartheta}$ é substancialmente maior do que a variância de $\hat{\vartheta}$. Então atentar-se à possível inequação $EQM(\tilde{\vartheta}) > EQM(\hat{\vartheta})$ é crucial quando usar a correção.

Uma aproximação do *intervalo de confiança bootstrap via percentis* com nível de confiança $100(1 - \alpha)\%$ é trivial. Sejam $\hat{\vartheta}_I^*$ e $\hat{\vartheta}_S^*$ os percentis $100(\alpha/2)$ e $100(1 - \alpha/2)$ da distribuição *bootstrap* de $\hat{\vartheta}$. Então o intervalo de $100(1 - \alpha)\%$ de confiança para ϑ é dado por

$$IC[\vartheta; 100(1 - \alpha)\%] \approx [\hat{\vartheta}_I^*; \hat{\vartheta}_S^*].$$

Entretanto, a cobertura real do intervalo definido acima é menor do que o nível nominal $1 - \alpha$ estipulado quando n é pequeno. Apesar de sua cobertura real ser próxima do nível nominal para n suficientemente grande, é possível atribuir um ajuste aos percentis para a criação de um intervalo de confiança melhor. Efron e Tibshirani (1993) definem esse novo intervalo como nome *intervalo de confiança bootstrap via percentis acelerados e não-viesados* e detalham esse assunto em seu trabalho.

A implementação do *bootstrap* é feita da seguinte maneira. A partir do conjunto de dados originais (amostra), retira-se uma amostra de tamanho B com reposição (amostra *bootstrap* ou reamostragem), onde B é um número enorme, geralmente entre 1000 a 10000 dependendo do poder computacional. Para cada reamostragem $b = 1, \dots, B$ calculam-se as estatísticas do teste $X_{M(b)}^2$ para o teste de independência marginal múltipla (MMI) e $X_{S(b)}^2$ para o teste de independência marginal em pares simultâneos (SPMI). O valor-p é igual a proporção de estatísticas $X_{M(b)}^2$ maior ou igual ao valor da estatística X_M^2 para o teste MMI e a proporção de estatísticas $X_{S(b)}^2$ maior ou igual ao valor da estatística X_S^2 , ou seja,

$$\frac{1}{B} \sum_{b=1}^B I(X_{M(b)}^2 \geq X_M^2) \text{ e } \frac{1}{B} \sum_{b=1}^B I(X_{S(b)}^2 \geq X_S^2),$$

onde $I(\cdot)$ é uma função indicadora. Rejeitam-se as hipóteses de MMI e SPMI para valores-p ínfimos. O teste de SPMI rejeita a um nível de significância nominal e real α quando SPMI é verdadeiro, como visto em Bilder e Loughin (2004).

As vantagens do *bootstrap* são a sua simplicidade, suas estimativas diretas da característica de interesse e consistente assintoticamente. No entanto, uma desvantagem são as condições para garantir a consistência assintótica como independência entre amostras *bootstrap*. Outra desvantagem é estudada por Athreya (1987), onde alerta o uso do *bootstrap* tradicional caso a população tenha distribuição de cauda pesada, não tenha variância finita ou não apresente quarto momento, pois a convergência da distribuição *bootstrap* não convergirá para o limite da característica de interesse.

2.2.3 Correções de Rao-Scott

Um dos propósitos das correções de Rao-Scott (1981) é o aperfeiçoamento da estatística de teste X_S^2 de Pearson tradicional em análises de independência de dados com delineamentos complexos. O emprego de estratificação ou *clusters* em amostragens não satisfaz a suposição de independência necessária e Rao e Scott propõem novas medidas. A

aproximação da distribuição de probabilidade assintótica de X_S^2 pode ser encontrada com essas correções.

A correção de primeira ordem ajusta a expectativa de X_S^2 para se tornar igual a de uma variável aleatória χ_{IJ}^2 . Contudo, essa correção não é suficiente para uma aproximação da distribuição de probabilidade de X_S^2 por uma χ_{IJ}^2 , causando rejeição demasiada frequente da hipótese nula H_0 quando SPMI é verdadeira. Por esse motivo, a correção de segunda ordem é a alternativa. Além disso, a estatística do teste ajustada não é invariante a designação de 0 ou 1 como a resposta positiva. Dessa forma, a correção de segunda ordem de Rao-Scott é proposta como alternativa para testar independência marginal em pares simultaneamente (SPMI).

Para testar SPMI, Bilder e Loughin (2004) definem λ_k como os coeficientes da aproximação de distribuição para X_S^2 para $k = 1, \dots, IJ$. A distribuição da estatística X_S^2 corrigida por um ajuste de segunda ordem de Rao-Scott é

$$X_{RS2}^2 = \frac{IJ \cdot X_S^2}{\sum_{k=1}^{IJ} \lambda_k^2},$$

podendo ser aproximada por uma variável aleatória χ^2 com $(IJ)^2 / \sum_{k=1}^{IJ} \lambda_k^2$ graus de liberdade, onde o número de graus de liberdade pode não ser um valor inteiro. Valores de X_{RS2}^2 maiores do que o quantil $1 - \alpha$ dessa distribuição χ^2 indicam evidências contra SPMI. Bilder e Loughin (2004) mostram em detalhes o quão bom é esse procedimento de teste, mas pode ser um pouco conservador.

2.3 Modelo log-linear marginal generalizado

Além de testes de hipótese para independência marginal, outra abordagem viável é o uso de modelos log-lineares para verificar a associação entre variáveis. Um modelo log-linear é um caso específico de modelos lineares generalizados para dados advindos de contagem com distribuição de probabilidade de Poisson cujas variáveis não apresentam distinção entre variáveis dependentes ou independentes, sendo todas elas tratadas como variáveis resposta. Por conta da natureza dos dados CATA, Bilder e Loughin (2007, 2009) propõem o modelo log-linear marginal generalizado que é um modelo log-linear que considera a estrutura de associação entre MRCVs.

Antes de introduzir o modelo log-linear marginal generalizado, uma breve revisão acerca de modelos log-lineares usuais será feita. Considere as representações de tabelas de contagem como na Figura 3 para cada nível $i = 1, \dots, I$ de X_i e $j = 1, \dots, J$ de Y_j . Note que,

$$\mu_{ij} = m_{++}\pi_{ij}; \quad \mu_{i+} = m_{++}\pi_{i+}; \quad \mu_{+j} = m_{++}\pi_{+j} \text{ e}; \quad \mu_{++} = m_{++}\pi_{++} = m_{++}.$$

Figura 2 – Formas gerais de tabelas de contingência para contagens, probabilidades e frequências esperadas.

		Y			T
		1	...	J	
X	I	m_{11}	...	m_{1j}	m_{1+}
	⋮	⋮	⋱	⋮	⋮
	I	m_{i1}	...	m_{ij}	m_{I+}
T		m_{+1}	...	m_{+J}	m_{++}

		Y			T
		1	...	J	
X	I	π_{11}	...	π_{1j}	π_{1+}
	⋮	⋮	⋱	⋮	⋮
	I	π_{i1}	...	π_{ij}	π_{I+}
T		π_{+1}	...	π_{+J}	π_{++}

		Y			T
		1	...	J	
X	I	μ_{11}	...	μ_{1j}	μ_{1+}
	⋮	⋮	⋱	⋮	⋮
	I	μ_{i1}	...	μ_{ij}	μ_{I+}
T		μ_{+1}	...	μ_{+J}	μ_{++}

Fonte: Elaboração do autor.

Sob independência,

$$\mu_{ij} = m_{++}\pi_{i+}\pi_{+j} = \frac{\mu_{i+}\mu_{+j}}{\mu_{++}}.$$

Quando as contagens em uma tabela de contingência seguem uma distribuição de probabilidade Poisson e as variáveis explicativas são todas categóricas, o uso de modelos log-lineares são comuns. Seu propósito é descrever a relação entre as variáveis categóricas e sua estrutura de interação. Representam-se tais modelos pela forma geral aditiva

$$\log(\mu_{ij}) = \log\left(\frac{\mu_{i+}\mu_{+j}}{\mu_{++}}\right) = \log(\mu_{i+}) + \log(\mu_{+j}) - \log(\mu_{++}).$$

Referente à uma tabela de contingência 2×2 , um modelo de independência para o modelo log-linear pode ser representado na tipologia ANOVA como

$$Y_{ij} = \mu + \alpha_i + \beta_j,$$

onde

$$Y_{ij} = \log(\mu_{ij}); \mu = -\log(\mu_{++}); \alpha_i = \log(\mu_{i+}) \text{ e; } \beta_j = \log(\mu_{+j}).$$

No entanto, a notação usual na literatura acerca de modelos log-lineares é descrita como a seguir:

$$\log(\mu_{ij}) = \lambda + \lambda_i^X + \lambda_j^Y,$$

onde λ é o efeito geral que controla o tamanho da amostra e ajusta uma constante, λ_i^X é o efeito marginal da variável X e λ_j^Y é o efeito marginal da variável Y .

O número de parâmetros do modelo log-linear é $I+J+1$, sendo $(I-1)(J-1)$ destes parâmetros estimáveis. Como consequência desse fato, temos um número de parâmetros maior do que um número de equações para a estimação, implicando na necessidade de restrições para garantir sua estimabilidade. Duas soluções para esse impasse são possíveis:

- $\sum_{i=1}^I \lambda_i^X = 0$ e; $\sum_{j=1}^J \lambda_j^Y = 0$;

- $\lambda_i^X = 0$ e; $\lambda_j^Y = 0$ para algum $i = 1, \dots, I$ e $j = 1, \dots, J$ arbitrários.

O modelo saturado considera variáveis estatisticamente dependentes, satisfazendo

$$\log(\mu_{ij}) = \lambda + \lambda_i^X + \lambda_j^Y + \lambda_{ij}^{XY},$$

onde λ_{ij}^{XY} são termos de associação indicadores de desvios de independência com

$$(I - 1) + (J - 1) + (I - 1)(J - 1) + 1 = IJ \text{ graus de liberdade.}$$

Analogamente para tabelas tridimensionais ($I \times J \times K$), os possíveis modelos que permitem termos de dependência entre variáveis são descritos a seguir.

- **Associação conjunta - modelo saturado** $[XYZ]$: Todos os efeitos individuais e as possíveis interações das variáveis são considerados.

$$\log(\mu_{ijk}) = \lambda + \lambda_i^X + \lambda_j^Y + \lambda_k^Z + \lambda_{ij}^{XY} + \lambda_{ik}^{XZ} + \lambda_{jk}^{YZ} + \lambda_{ijk}^{XYZ}.$$

Este modelo tem JKK parâmetros estimáveis. Por apresentar todas as interações possíveis bem como os efeitos individuais, não tem graus de liberdade. A razão de chances de duas variáveis condicionadas a uma outra não necessariamente é igual a 1.

- **Associação homogênea** $[XY, XZ, YZ]$: Considera os efeitos individuais e possíveis interações das variáveis duas a duas, mas desconsidera a interação das três variáveis conjuntamente.

$$\log(\mu_{ijk}) = \lambda + \lambda_i^X + \lambda_j^Y + \lambda_k^Z + \lambda_{ij}^{XY} + \lambda_{ik}^{XZ} + \lambda_{jk}^{YZ}.$$

Apresenta

$$\begin{aligned} g.l. &= IJK - 1 - (I - 1) - (J - 1) - (K - 1) - (I - 1)(J - 1) \\ &\quad - (I - 1)(K - 1) - (J - 1)(K - 1) \\ &= (I - 1)(J - 1)(K - 1) \end{aligned}$$

graus de liberdade e razões de chances não necessariamente iguais a 1 devido a possíveis interações significantes, indicando dependência nesses casos.

- **Independência condicional** $[XY, XZ]$, $[XY, YZ]$ ou $[XZ, YZ]$: Os efeitos individuais ainda continuam no modelo e os efeitos de interesse são os representados pela notação. Exemplificando, escrever o modelo de independência condicional $[XY, XZ]$ é equivalente a

$$\log(\mu_{ijk}) = \lambda + \lambda_i^X + \lambda_j^Y + \lambda_k^Z + \lambda_{ij}^{XY} + \lambda_{ik}^{XZ}$$

com $IJK - 1 - (I - 1) - (J - 1) - (K - 1) - (I - 1)(J - 1) - (I - 1)(K - 1) = I(J - 1)(K - 1)$ graus de liberdade. A razão de chances, nesse caso, para YZ dado X $\theta_{YZ(i)} = 1$ pela independência condicional estabelecida. Da mesma forma, para os outros modelos de independência condicional, $\theta_{XZ(j)} = 1$ e $\theta_{XY(z)} = 1$ para os modelos $[XY, YZ]$ e $[XZ, YZ]$ respectivamente.

- **Independência conjunta** $[XY, Z], [XZ, Y], [YZ, X]$: Os efeitos individuais são considerados bem como a interação de interesse descrita pela notação. Tomando como exemplo $[XZ, Y]$, modela-se:

$$\log(\mu_{ijk}) = \lambda + \lambda_i^X + \lambda_j^Y + \lambda_k^Z + \lambda_{ik}^{XZ}.$$

Possui $IJK - 1 - (I - 1) - (J - 1) - (K - 1) - (I - 1)(K - 1) = (IK - 1)(J - 1)$ graus de liberdade. Como cada conjunto XY dado Z e YZ dado X é independente, as razões de chances condicionais $\theta_{XY(k)}$ e $\theta_{YZ(i)}$ são iguais a 1. Analogamente para os outros modelos de independência conjunta, para $[XY, Z]$ as razões de chance de XZ dado Y $\theta_{XZ(j)} = 1$ e de YZ dado X $\theta_{YZ(i)} = 1$ e para $[YZ, X]$ as razões de chance de XY dado Z $\theta_{XY(k)} = 1$ e XZ dado Y $\theta_{XZ(j)} = 1$.

- **Independência completa** $[X, Y, Z]$: Apenas os efeitos individuais são significantes:

$$\log(\mu_{ijk}) = \lambda + \lambda_i^X + \lambda_j^Y + \lambda_k^Z.$$

Apresenta $IJK - 1 - (I - 1) - (J - 1) - (K - 1)$ graus de liberdade e razões de chance parciais iguais a 1 devido a natureza da independência.

Esses modelos são hierárquicos, ou seja, a consideração de interações em um modelo requer a introdução de efeitos individuais relativos a essas interações.

Para testar qualidade desses ajustes, usa-se das hipóteses:

$$\begin{cases} H_0 : \text{O modelo é adequado} \\ H_1 : \text{O modelo é inadequado} \end{cases}$$

e pode-se usar as estatísticas

$$Deviance = 2 \sum_i \sum_j \sum_k n_{ijk} \log \left(\frac{n_{ijk}}{\hat{\mu}_{ijk}} \right) \text{ e}$$

$$X^2 = \sum_i \sum_j \sum_k \frac{(n_{ijk} - \hat{\mu}_{ijk})^2}{\hat{\mu}_{ijk}},$$

onde razões elevadas para $Deviance/g.l.$ ou $X^2/g.l.$ indicam rejeição da hipótese de um modelo adequado.

Visualmente, um modelo adequado apresenta resíduos de Pearson

$$e_{ijk} = \frac{(n_{ijk} - \hat{\mu}_{ijk})^2}{\sqrt{\hat{\mu}_{ijk}}}$$

ou resíduos ajustados

$$\frac{e_{ijk}}{1 - h_{ijk}}$$

distribuídos aproximadamente por uma $N(0, 1)$, onde h_{ijk} é ponto de alavanca da célula (i, j, k) .

Sejam \mathfrak{P} uma população de tamanho N , \mathfrak{s} uma amostra selecionada de \mathfrak{P} de tamanho n seguindo um certo plano amostral e $p_w, w = 1, \dots, N$ os pesos de pesquisa com o intuito de encontrar estimativas não viesadas da população. Considere, para fins de explicação, o caso com duas MRCVs como no exemplo de uso de tabaco e problemas respiratórios. Sejam $\mathbf{X} = (X_1, \dots, X_I)$ e $\mathbf{Y} = (Y_1, \dots, Y_J)$ os vetores de respostas binárias, onde X_i e Y_j são as respostas binárias $\{0, 1\}$ para os itens i e j dos vetores \mathbf{X} e \mathbf{Y} . Sejam também $\mathbf{x}_w = (x_{w1}, \dots, x_{wI})$ e $\mathbf{y}_w = (y_{w1}, \dots, y_{wJ})$ para $w = 1, \dots, N$ os valores correspondentes observados na população.

Defina a contagem marginal da população para $(x_i = u, y_j = v)$ como

$$m_{uv(ij)} = \sum_{w \in \mathfrak{P}} I(x_{wi} = u, y_{wj} = v),$$

onde $I(\cdot)$ é uma função indicadora e $u, v \in \{0, 1\}$. Ilustrando para a tabela de contingência 2×2 para a combinação de itens $(i = 1, j = 1)$, a contagem marginal da população para uma resposta $(x_i = 1, y_j = 1)$ para essa tabela é dada pela soma do número de indivíduos que marcaram essa resposta positiva nessa tabela, para uma resposta $(x_i = 0, y_j = 1)$ sua contagem marginal seria a soma do número de indivíduos para essa resposta e analogamente para as respostas $(x_i = 1, y_j = 0)$ e $(x_i = 0, y_j = 0)$. O processo descrito para a tabela $(i = 1, j = 1)$ anterior é feito para todas as combinações (i, j) com $i = 1, \dots, I$ e $j = 1, \dots, J$, resultando em contagens marginais para todas as IJ tabelas 2×2 . A tabela a seguir ilustra a forma geral das contagens marginais da população.

Tabela 5 – Forma geral de uma tabela de resposta ao item

			Y				
			1		2		...
			1	0	1	0	...
X	1	1	$m_{11(11)}$	$m_{10(11)}$	$m_{11(12)}$	$m_{10(12)}$...
		0	$m_{01(11)}$	$m_{00(11)}$	$m_{01(12)}$	$m_{00(12)}$...
	2	1	$m_{11(21)}$	$m_{10(21)}$	$m_{11(22)}$	$m_{10(22)}$...
		0	$m_{01(21)}$	$m_{00(21)}$	$m_{01(22)}$	$m_{00(22)}$...
⋮	⋮	⋮	⋮	⋮	⋮	⋮	

Fonte: Elaboração do autor.

O estimador de $m_{uv(ij)}$ é

$$\widehat{m}_{uv(ij)} = \mu_{uv(ij)} = \sum_{w \in \mathfrak{s}} p_w \times I(x_{wi} = u, y_{wj} = v).$$

A proporção marginal populacional com pesos de pesquisa é dado por

$$\pi_{uv(ij)} = \frac{m_{uv(ij)}}{N}$$

e seu estimador é

$$\widehat{\pi}_{uv(ij)} = p_{uv(ij)} = \frac{\mu_{uv(ij)}}{\widehat{N}},$$

onde $\widehat{N} = \sum_{w \in \mathfrak{s}} p_w$.

Substituindo a notação da Tabela 5 por $\widehat{m}_{uv(ij)} = \mu_{uv(ij)}$ ou por $\widehat{\pi}_{uv(ij)} = p_{uv(ij)}$ temos a tabela de resposta ao item para as frequências esperadas e probabilidades para cada combinação de item, respectivamente.

O modelo log-linear para as contagens marginais totais \mathbf{m} relaciona o logaritmo dessas contagens aos parâmetros capazes de resumir a estrutura associativa entre duas MRCVs, sendo um aprimoramento do modelo log-linear de independência numa tabela de contingência 2×2 tradicional. O modelo log-linear marginal generalizado providencia parâmetros para todas as IJ tabelas 2×2 marginais simultaneamente por

$$\log(m_{uv(ij)}) = \lambda_{ij} + \lambda_{u(ij)}^X + \lambda_{v(ij)}^Y, \quad i = 1, \dots, I, j = 1, \dots, J, u = 0, 1, v = 0, 1,$$

onde λ_{ij} , $\lambda_{u(ij)}^X$ e $\lambda_{v(ij)}^Y$ são parâmetros desconhecidos acerca das contagens totais em cada sub-tabela, dos totais marginais nas linhas e dos totais marginais nas colunas, respectivamente, em cada sub-tabela (i, j) , $i = 1, \dots, I$, $j = 1, \dots, J$. Essa estrutura é definida também denominada SPMI por conta da suposição de independência entre cada item \mathbf{X} e \mathbf{Y} advinda de seus parâmetros apresentarem contagens marginais em cada sub-tabela.

Da mesma forma que nos modelos log-lineares usuais, parâmetros adicionais permissores de diferentes associações e independências entre itens de \mathbf{X} e \mathbf{Y} podem ser incluídos. Adicionar um termo constante λ_{uv} para todo i e j resulta num modelo parecido com o de independência conjunta. Adicionando o termo $\lambda_{uv(i)}^X$ ou $\lambda_{uv(j)}^Y$ no modelo anterior, obtém-se modelos semelhantes aos de independência condicional, permitindo variação do logaritmo da razão de chances ao longo dos itens de \mathbf{X} ou \mathbf{Y} , respectivamente. Da mesma forma, acrescentando ambos os termos, varia-se o logaritmo das contagens tanto em \mathbf{X} como em \mathbf{Y} , resultando em um modelo similar ao de associação homogênea. Por fim, o modelo saturado resultante da adição de $\lambda_{uv(ij)}^{XY}$ no modelo de independência conjunta anterior equivale à análise alternativa de examinar as IJ sub-tabelas independentemente, um modelo de associação conjunta análogo aos modelos log-lineares usuais.

A parametrização pode ser generalizada matricialmente por

$$\log(\mathbf{m}_{uv(ij)}) = \mathbf{x}_{uv(ij)}\boldsymbol{\beta} = \mathbf{X}\boldsymbol{\beta},$$

onde $\boldsymbol{\beta}$ é um vetor $r \times 1$ de parâmetros e $\mathbf{x}_{uv(ij)}$ é um vetor $1 \times r$ de 0's e 1's adequado e com correspondência ao modelo de interesse. A generalização matricial mais à direita da igualdade acima é uma reescrita onde $\mathbf{X} = (\mathbf{x}'_{00(11)} \parallel \mathbf{x}'_{01(11)} \parallel \dots \parallel \mathbf{x}'_{11(IJ)})$ com \parallel denotando concatenação horizontal. Mesmo ajustado diretamente aos totais marginais populacionais estimados $\boldsymbol{\mu}$ como contagens multinomiais, essa errônea suposição é corrigida para utilizar procedimentos inferenciais. As equações estimadoras de pseudo verossimilhança resolvidas para produzir as contagens preditas são $\mathbf{X}'\boldsymbol{\mu} = \mathbf{X}'\tilde{\boldsymbol{\mu}}$, onde $\tilde{\boldsymbol{\mu}}$ é um vetor $4IJ \times 1$ de contagens de células preditas com valores $\tilde{\mu}_{uv(ij)} = \exp\{\mathbf{x}_{uv(ij)}\hat{\boldsymbol{\beta}}\}$ e $\hat{\boldsymbol{\beta}}$ é o vetor de parâmetros estimado. O ajuste deste modelo é geralmente robusto para dados esparsos entre combinação de repostas (\mathbf{x}, \mathbf{y}) .

Defina $\mu_{uv(ij)}^{(0)}$ e $\mu_{uv(ij)}^{(A)}$ os totais populacionais preditos sob hipóteses nula e alternativa quaisquer respectivamente, com a hipótese nula aninhada à hipótese alternativa e defina também $p_{uv(ij)}^{(0)}$ e $p_{uv(ij)}^{(A)}$ as proporções populacionais resultantes do ajuste dos respectivos modelos anteriores. A estatística de Pearson para comparação dos modelos é

$$X^2 = n \sum_{i,j,u,v} \frac{(p_{uv(ij)}^{(A)} - p_{uv(ij)}^{(0)})^2}{p_{uv(ij)}^{(0)}} = \frac{n}{N} \sum_{i,j,u,v} \frac{(\mu_{uv(ij)}^{(A)} - \mu_{uv(ij)}^{(0)})^2}{\mu_{uv(ij)}^{(0)}}.$$

Autores anteriores como Agresti e Liu (1999) estabeleceram que mesmo sob amostragem aleatória simples as contagens marginais de uma tabela de resposta ao item não apresenta distribuição multinomial por múltiplas razões já discorridos na subseção 2.2. Bilder e Loughin (2009) demonstram que a distribuição assintótica de X^2 é uma combinação linear de variáveis aleatórias com distribuição χ_1^2 , acatando o delineamento da pesquisa e as respostas múltiplas. Rao e Scott (1984) propõem correções de primeira e segunda ordem para estatísticas similares a X^2 a fim de aproximá-la a uma distribuição Qui-Quadrado na análise de associação entre variáveis categóricas de única resposta em delineamentos complexos de amostragem. Thomas, Singh e Roberts (1996) mostram em detalhe que uma distribuição F aproximada de versões modificadas de X^2 permite um nível real mais próximo do nível nominal do teste.

O resíduo de Pearson padronizado é dado por

$$r_{uv(ij)} = \frac{(\mu_{uv(ij)} - \tilde{\mu}_{uv(ij)})}{\sqrt{\hat{V}}},$$

onde \hat{V} é a variância assintótica estimada de $\mu_{uv(ij)} - \tilde{\mu}_{uv(ij)}$, pode ser usado como medida de desvio de um modelo adequado, usando o elemento diagonal correspondente da matriz de covariância assintótica dos resíduos.

3 Aplicação

O questionário de Percepção sobre Computação do projeto Meninas da Computação foi respondido por 3707 pessoas durante os anos de 2011 a 2014 e 440 pessoas em 2019, totalizando 4147 respostas coletadas. 212 respostas foram consideradas inválidas devido à marcações no campo Masculino ou não marcação para a variável Sexo e/ou não marcação na variável Área de Interesse. Dessa forma, os dados finais utilizados apresentam 3530 respostas cobrindo os anos de 2011 a 2014 e 405 respostas de 2019, resultando em 3935 respostas válidas.

Tabela 6 – Respostas para o questionário Percepção sobre Computação do projeto Meninas na Computação

Período	Respostas	Respostas inválidas	Respostas válidas
2011-2014	3707	177	3530
2019	440	35	405
Total	4147	212	3935

Fonte: Elaboração do autor.

A variável de resposta singular Área de Interesse é aliciante para a análise e apresenta os níveis Exatas, Biológicas e Saúde e Humanas codificados em Exat., Bio. & S. e Hum. respectivamente. As duas variáveis categóricas de resposta múltipla do questionário se referem aos locais de uso de computador e atividades realizadas no computador. Os locais de uso de computador são respectivamente: em casa; em casa de parentes; em casa de amigos; na escola; no trabalho; em *lan-house*; biblioteca e; centros de inclusão digital. As atividades realizadas no computador são respectivamente: edição de texto (*Word*, outros); edição de imagem; planilha (*Excel*, outros); banco de dados (*Access*, outros); acesso a *internet* (pesquisas sobre conteúdos, notícias, outros); redes sociais (*Facebook*, *Orkut*, etc); *e-mail*; jogos; desenvolvimentos de páginas; programação e; outras atividades.

Hipóteses a serem investigadas são a existência de influência dos locais de uso do computador e das atividades realizadas no computador na área de interesse das respondentes, presença pertinente de uma dependência entre determinadas atividades e o local de uso do computador e se houve mudança significativa dessas dependências entre os anos de 2011 a 2014 com 2019.

Começamos a análise pela visualização de tabelas de contingência marginal e tabelas de resposta ao item. Nas tabelas deste último tipo, as codificações 1 e 0 significam marcações positivas e negativas, respectivamente, de cada respondente no questionário para cada possível local de uso do computador ou atividade realizada no computador. Além disso, gráficos de mosaico também são usados para verificar se há alguma relação

aparente entre a variável Área de Interesse com as MRCVs Local de Uso do Computador e Atividades Realizadas no Computador. Um gráfico de mosaico dispõe as contagens observadas das células de uma tabela de contingência em blocos cuja área é proporcional a essas contagens, podendo ser usadas até quatro tabelas de contingência na construção do gráfico.

O pacote *MRCV* criado por Bilder e Koziol (2014) para o *software* R fora utilizado a fim de produzir as tabelas de contingência marginal e de resposta ao item, conduzir os testes de independência marginal múltipla e de independência marginal em pares simultâneos e ajustar modelos para os dados. O pacote apresenta funções para comparar modelos e analisar seus resíduos levando em consideração a dependência entre as respostas.

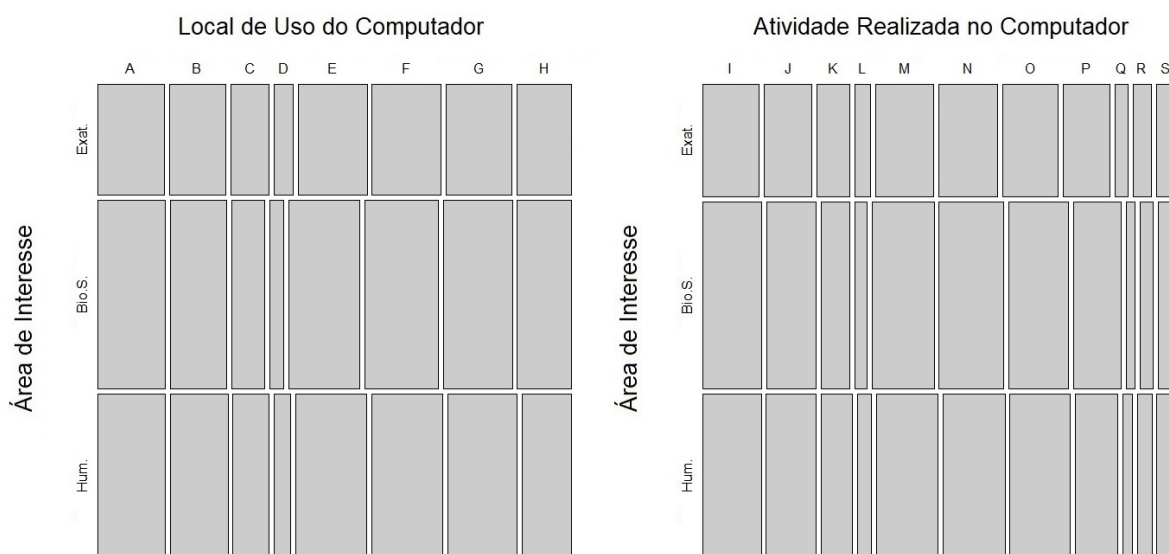
3.1 Respostas de 2011 a 2014

Durante o período de 2011 a 2014, a Tabela 7 indica que o uso do computador no trabalho e em *lan-house* parece ter uma proporção maior de usuárias interessadas nas áreas de Humanas e Biológicas e Saúde, respectivamente. O comportamento geral das proporções para os outros locais de uso do computador parece não ser muito influenciado pela área de interesse das respondentes. A proporção de uso do computador em casa, em casa de parentes, em casa de amigos, na escola, na biblioteca ou em centros de inclusão digital parece ser semelhante para as três áreas de interesse.

Para o mesmo período, a Tabela 8 indica que interessadas em exatas parecem usar planilhas, acessar banco de dados, desenvolver páginas e programar significativamente mais frequente do que as meninas de outras áreas. No entanto, parece que editar textos ou imagens sejam atividades realizadas mais frequentemente por quem não tem exatas como área de interesse. Não parece existir relação entre a área de interesse com edição de imagens ou uso de redes sociais.

Na Figura 3 estão apresentados os gráficos de mosaico para Área de Interesse vs. Local de Uso do Computador e Área de Interesse vs. Atividade Realizada no Computador relativos às Tabelas 7 e 8. As contagens para o uso do computador no trabalho e em *lan-house* parecem ser proporcionalmente maiores nas áreas de Humanas e Biológicas e Saúde, respectivamente, sendo um indício do gráfico de mosaico semelhante à análise da Tabela 7. Da mesma forma, a frequência do uso do computador pelas respondentes interessadas em Exatas para as atividades planilhas, acessar banco de dados, desenvolver páginas e programar ou editar textos e imagens para interessadas nas outras duas áreas é uma interpretação do gráfico de mosaico semelhante à análise da Tabela 8.

Figura 3 – Gráficos de mosaico para as variáveis Área de Interesse vs. Local de Uso de Computador e Área de Interesse vs. Atividade Realizada no Computador (2011 - 2014).



Fonte: Elaboração do autor. A: em casa; B: em casa de parentes; C: em casa de amigos; D: na escola; E: no trabalho; F: em *lan-house*; G: biblioteca; H: centros de inclusão digital; I: edição de texto; J: edição de imagem; K: planilha; L: banco de dados; M: acesso a *internet*; N: redes sociais; O: *e-mail*; P: jogos; Q: desenvolvimentos de páginas; R: programação e; S: outras atividades.

Tabela 7 – Tabela de contingência marginal para as variáveis Área de Interesse e Local de Uso de Computador (2011 - 2014).

Local de Uso do Computador	Área de Interesse		
	Exatas	Biológicas e Saúde	Humanas
Em casa	754 (94,13%)	1361 (92,02%)	1164 (93,12%)
Em casa de parentes	536 (66,92%)	1009 (68,22%)	809 (64,72%)
Em casa de amigos	525 (65,54%)	948 (64,10%)	823 (65,84%)
Na escola	399 (49,81%)	666 (45,03%)	585 (46,80%)
No trabalho	151 (18,85%)	235 (15,89%)	322 (25,76%)
Em <i>lan-house</i>	256 (31,96%)	553 (37,39%)	377 (30,16%)
Biblioteca	215 (26,84%)	365 (24,68%)	307 (24,56%)
Centros de inclusão digital	58 (7,24%)	91 (6,15%)	93 (7,44%)
Total de respondentes (m_{ij})	801 (100,00%)	1479 (100,00%)	1250 (100,00%)

Fonte: Elaboração do autor. Porcentagem das respostas em relação a m_{ij} em parênteses.

Tabela 8 – Tabela de contingência marginal para as variáveis Área de Interesse e Atividade Realizada no Computador (2011 - 2014).

Atividade Realizada no Computador	Área de Interesse		
	Exatas	Biológicas e Saúde	Humanas
Edição de texto	716 (89,39%)	1239 (83,77%)	1090 (87,20%)
Edição de imagem	607 (75,78%)	1038 (70,18%)	927 (74,16%)
Planilha	413 (51,56%)	605 (40,91%)	578 (46,24%)
Banco de dados	207 (25,84%)	258 (17,44%)	265 (21,20%)
Acesso a <i>internet</i>	736 (91,89%)	1306 (88,30%)	1144 (91,52%)
Redes sociais	740 (92,38%)	1352 (91,41%)	1145 (91,60%)
<i>e-mail</i>	710 (88,64%)	1259 (85,13%)	1114 (89,12%)
Jogos	586 (73,16%)	1000 (67,61%)	801 (64,08%)
Desenvolvimento de páginas	169 (21,10%)	190 (12,85%)	180 (14,40%)
Programação	230 (28,71%)	281 (19,00%)	262 (20,96%)
Outros	261 (32,58%)	388 (26,23%)	370 (29,60%)
Total de respondentes (m_{ij})	801 (100,00%)	1479 (100,00%)	1250 (100,00%)

Fonte: Elaboração do autor. Porcentagem das respostas em relação a m_{ij} em parênteses.

A condução de testes de independência marginal múltipla (MMI) para a variável Área de Interesse com cada MRCV Local de Uso do Computador e Atividades Realizadas no Computador e independência marginal em pares simultâneos (SPMI) para as duas MRCVs indicam dependência estatisticamente significativa de pelo menos uma área de interesse com um local de uso do computador e uma das atividades para todos os métodos utilizados e dependência entre pelo menos um local e uma atividade. Pelas Tabelas 9 e 10, usar o computador no trabalho ou em *lan-house* e usar o computador para editar textos, usar planilhas e banco de dados, jogar jogos, desenvolver páginas e programar têm relação com a área de interesse das respondentes no período de 2011 a 2014. Usar o computador nos demais locais e usá-lo para editar imagens ou acessar redes sociais não têm relação com a área de interesse das respondentes para o mesmo período. Como os valores-p para acesso a *internet*, *e-mail* e outras atividades não são desprezíveis o suficiente, não é viável concluir algo sobre a relação dessas atividades com a área de interesse das respondentes. A Tabela 11 apresenta os valores-p para cada procedimento ao testar MMI e SPMI.

Tabela 9 – Saída R com as estatísticas Qui-Quadrado de Pearson não ajustadas e valores-p ajustados por Bonferroni para as variáveis Área de Interesse vs. Local de Uso do Computador (2011-2014).

Local de Uso do Computador	Estatística Qui-Quadrado de Pearson não ajustada ($X_{S(ij)}^2$)	Valor-p ajustado por Bonferroni (p_{ij})
Em casa	3,66	1,0000
Em casa de parentes	3,76	1,0000
Em casa de amigos	1,02	1,0000
Na escola	4,78	0,7344
No trabalho	42,11	0,0000
Em <i>lan-house</i>	17,12	0,0015
Biblioteca	1,62	1,0000
Centros de inclusão digital	2,00	1,0000

Fonte: Elaboração do autor.

Tabela 10 – Saída R com as estatísticas Qui-Quadrado de Pearson não ajustadas e valores-p ajustados por Bonferroni para as variáveis Área de Interesse vs. Atividade Realizada no Computador (2011-2014).

Atividade Realizada no Computador	Estatística Qui-Quadrado de Pearson não ajustada ($X_{S(ij)}^2$)	Valor-p ajustado por Bonferroni (p_{ij})
Edição de texto	15,27	0,0053
Edição de imagem	9,88	0,0785
Planilha	24,64	0,0000
Banco de dados	22,66	0,0001
Acesso a <i>internet</i>	11,1	0,0427
Redes sociais	0,67	1,0000
e-mail	11,36	0,0375
Jogos	18,38	0,0011
Desenvolvimento de páginas	28,48	0,0000
Programação	29,67	0,0000
Outros	10,71	0,0519

Fonte: Elaboração do autor.

Tabela 11 – Valores-p para os testes MMI e SPMI (2011-2014).

Método	AdI		Local vs. Atividade			
	vs. Local	vs. Atividade	Geral	Exatas	Biol.S.	Humanas
Bonferroni	< 0,0001	< 0,0001	< 0,0001	< 0,0001	< 0,0001	< 0,0001
X_S^2 <i>bootstrap</i>	< 0,0005	< 0,0005	< 0,0005	< 0,0005	< 0,0005	< 0,0005
Produto p_{ij} <i>bootstrap</i>	< 0,0005	< 0,0005	< 0,0005	< 0,0005	< 0,0005	< 0,0005
$\min_{ij} p_{ij}$ <i>bootstrap</i>	< 0,0005	< 0,0005	< 0,0005	< 0,0005	< 0,0005	< 0,0005
X_{RS2}^2	< 0,0001	< 0,0001	-	-	-	-

Fonte: Elaboração do autor. Área de Interesse abreviado por AdI e Biológicas e Saúde abreviado por Biol.S.

Tabela 12 – Valores-p por ajuste de Bonferroni para cada estatística $X_{S(i,j)}^2$ para o teste SPMI das variáveis Local de Uso de Computador e Atividade Realizada no Computador (2011 - 2014)

Local	Atividade										
	I	J	K	L	M	N	O	P	Q	R	S
A	0,000	0,000	0,003	1,000	0,000	0,000	0,000	1,000	1,000	0,079	1,000
B	0,000	0,000	1,000	0,168	0,000	0,000	0,000	0,000	0,076	0,001	0,000
C	0,000	0,000	1,000	0,254	0,000	0,000	0,000	0,000	0,001	0,000	0,000
D	0,000	0,000	0,000	0,000	0,011	0,002	0,000	0,000	0,000	0,000	0,000
E	0,000	0,001	0,000	0,000	0,006	1,000	0,000	1,000	0,000	1,000	0,043
F	1,000	1,000	1,000	0,030	1,000	1,000	1,000	0,000	0,268	1,000	0,000
G	0,000	0,000	0,000	0,000	0,122	1,000	0,000	0,000	0,000	0,000	0,000
H	0,147	1,000	0,000	0,000	1,000	1,000	1,000	0,078	0,000	0,000	0,000

Fonte: Elaboração do autor. A: em casa; B: em casa de parentes; C: em casa de amigos; D: na escola; E: no trabalho; F: em *lan-house*; G: biblioteca; H: centros de inclusão digital; I: edição de texto; J: edição de imagem; K: planilha; L: banco de dados; M: acesso a *internet*; N: redes sociais; O: *e-mail*; P: jogos; Q: desenvolvimentos de páginas; R: programação e; S: outras atividades.

Tabela 13 – Valores-p por ajuste de Bonferroni para cada estatística $X_{S(i,j)}^2$ para o teste SPMI das variáveis Local de Uso de Computador e Atividade Realizada no Computador para interessadas em Exatas (2011 - 2014)

Local	Atividade										
	I	J	K	L	M	N	O	P	Q	R	S
A	0,294	0,000	1,000	1,000	1,000	0,024	0,000	0,128	1,000	1,000	1,000
B	0,025	0,000	1,000	1,000	1,000	0,000	0,000	0,002	1,000	1,000	0,098
C	0,082	0,000	1,000	1,000	1,000	0,001	0,003	0,000	1,000	1,000	0,138
D	0,038	0,001	0,000	0,000	1,000	1,000	0,002	0,903	0,905	0,040	0,227
E	0,003	1,000	0,000	0,000	1,000	1,000	0,016	1,000	1,000	1,000	0,689
F	1,000	1,000	1,000	1,000	1,000	1,000	1,000	1,000	1,000	1,000	0,646
G	0,001	0,000	0,000	0,000	1,000	1,000	0,001	0,233	0,101	0,000	0,112
H	1,000	1,000	0,518	0,000	1,000	1,000	1,000	1,000	0,002	0,001	0,038

Fonte: Elaboração do autor. A: em casa; B: em casa de parentes; C: em casa de amigos; D: na escola; E: no trabalho; F: em *lan-house*; G: biblioteca; H: centros de inclusão digital; I: edição de texto; J: edição de imagem; K: planilha; L: banco de dados; M: acesso a *internet*; N: redes sociais; O: *e-mail*; P: jogos; Q: desenvolvimentos de páginas; R: programação e; S: outras atividades.

Tabela 14 – Valores-p por ajuste de Bonferroni para cada estatística $X_{S(i,j)}^2$ para o teste SPMI das variáveis Local de Uso de Computador e Atividade Realizada no Computador para interessadas em Biológicas e Saúde (2011 - 2014)

Local	Atividade										
	I	J	K	L	M	N	O	P	Q	R	S
A	0,000	0,000	0,253	1,000	0,000	0,000	0,000	1,000	1,000	1,000	1,000
B	0,002	0,000	1,000	1,000	0,272	0,003	0,001	0,000	1,000	0,214	0,603
C	0,000	0,000	1,000	0,445	0,002	0,000	0,000	0,000	0,163	0,085	0,001
D	0,000	0,000	0,004	1,000	0,317	1,000	0,000	0,000	0,071	1,000	0,000
E	0,020	0,436	0,000	0,000	0,969	1,000	0,006	1,000	0,068	1,000	1,000
F	1,000	1,000	1,000	0,499	1,000	1,000	1,000	0,000	1,000	1,000	0,135
G	0,003	0,000	0,322	0,005	1,000	1,000	0,014	0,244	0,013	0,222	0,000
H	1,000	1,000	0,008	0,000	1,000	1,000	1,000	1,000	0,000	0,000	0,000

Fonte: Elaboração do autor. A: em casa; B: em casa de parentes; C: em casa de amigos; D: na escola; E: no trabalho; F: em *lan-house*; G: biblioteca; H: centros de inclusão digital; I: edição de texto; J: edição de imagem; K: planilha; L: banco de dados; M: acesso a *internet*; N: redes sociais; O: *e-mail*; P: jogos; Q: desenvolvimentos de páginas; R: programação e; S: outras atividades.

Tabela 15 – Valores-p por ajuste de Bonferroni para cada estatística $X_{S(i,j)}^2$ para o teste SPMI das variáveis Local de Uso de Computador e Atividade Realizada no Computador para interessadas em Humanas (2011 - 2014)

Local	Atividade										
	I	J	K	L	M	N	O	P	Q	R	S
A	0,000	0,000	1,000	1,000	1,000	0,007	0,003	1,000	1,000	1,000	1,000
B	0,089	0,000	1,000	0,099	0,169	0,000	0,013	0,000	1,000	0,034	0,001
C	0,093	0,000	1,000	1,000	0,000	0,000	0,034	0,000	1,000	0,376	0,011
D	0,000	0,000	0,019	0,318	1,000	0,340	0,000	1,000	0,011	0,161	1,000
E	0,004	0,629	0,000	0,000	1,000	1,000	0,001	1,000	0,192	1,000	1,000
F	1,000	1,000	1,000	0,354	1,000	1,000	1,000	0,002	1,000	1,000	0,011
G	0,231	0,041	0,000	0,020	1,000	1,000	0,209	0,126	0,275	0,385	1,000
H	1,000	1,000	0,218	0,002	1,000	1,000	1,000	1,000	0,032	0,000	0,009

Fonte: Elaboração do autor. A: em casa; B: em casa de parentes; C: em casa de amigos; D: na escola; E: no trabalho; F: em *lan-house*; G: biblioteca; H: centros de inclusão digital; I: edição de texto; J: edição de imagem; K: planilha; L: banco de dados; M: acesso a *internet*; N: redes sociais; O: *e-mail*; P: jogos; Q: desenvolvimentos de páginas; R: programação e; S: outras atividades.

O valor-p *bootstrap*, mínimo dos valores-p *bootstrap* e produto dos valores-p *bootstrap* para 1999 amostras *bootstrap* são menores do que 0,0005 e os valores-p ajustado por Bonferroni e Segunda Ordem de Rao-Scott também são ínfimos, favorecendo rejeição da hipótese nula de inexistência de dependência entre as respostas das participantes no questionário.

O uso do computador no trabalho ou em *lan-house* tem relação com a área de interesse da respondente. Como suspeitado, o uso do computador no trabalho e em *lan-house* bem como usar editores de texto e planilhas, acessar banco de dados, jogar jogos

de computador, desenvolver páginas e programar têm uma relação de dependência com a área de interesse para respondentes de 2011 a 2014.

As Tabelas 12 a 15 apresentam os valores-p por ajuste de Bonferroni para cada estatística $X_{S(i,j)}^2$ na condução do teste SPMI para as variáveis Local de Uso do Computador e Atividade Realizada no Computador no período de 2011 a 2014 para todas as respostas e respostas das interessadas nas áreas de Exatas, Biológicas e Saúdes e Humanas, respectivamente. Células cinza-escuras indicam dependência significativa entre determinado local e determinada atividade, células com fundo branco indicam independência significativa entre o local e a atividade e células cinza-claras indicam incerteza quanto a dependência ou independência dos níveis das variáveis. Por conta de seus valores-p, não se deve concluir dependência ou independência pelo teste SPMI por ajuste de Bonferroni para as seguintes combinações: em geral, realizar outras atividades no trabalho e usar banco de dados em *lan-house*; editar texto e programar na escola e realizar outras atividades em centros de inclusão digital entre as interessadas em Exatas; desenvolver páginas no trabalho entre as interessadas em Biológicas e Saúde e; programar em casa de parentes, usar *e-mail* em casa de amigos, editar imagem na biblioteca e desenvolver páginas em centros de inclusão digital entre as interessadas em Humanas.

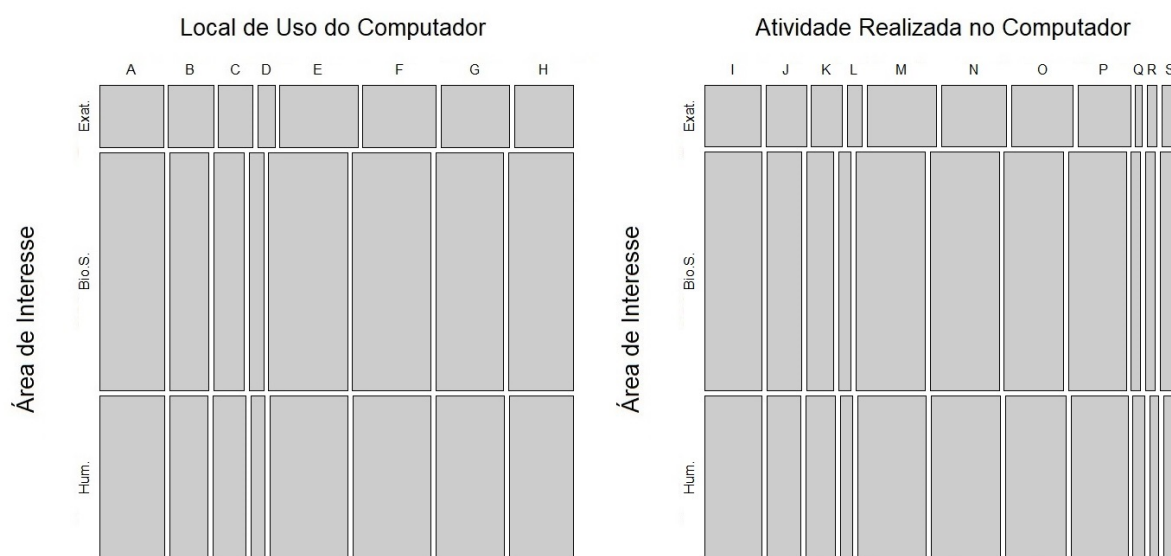
3.2 Respostas de 2019

Já para o período de 2019, a Tabela 16 indica variação de uso do computador das respondentes para diferentes áreas de interesse nos locais: em casa, em casa de parentes, em casa de amigos, na escola, no trabalho e em *lan-house*. Não parece haver diferença na frequência de uso do computador na biblioteca ou em centros de inclusão digital para diferentes áreas de interesse.

Durante o ano de 2019, a Tabela 17 indica que as atividades edição de imagem, uso de planilha e desenvolvimento de páginas parecem estar relacionadas com o interesse das respondentes em determinadas áreas de interesse: meninas interessadas em Exatas praticam edição de imagem e usam planilha mais que as demais meninas e meninas interessadas em Humanas desenvolvem páginas mais frequentemente do que as demais.

A Figura 4 apresenta os gráficos de mosaico para Área de Interesse vs. Local de Uso do Computador e Área de Interesse vs. Atividade Realizada no Computador relativos às Tabelas 16 e 17. As frequências de uso do computador em casa de parentes e de amigos aparentam ser proporcionalmente maiores na área de Exatas enquanto as respondentes interessadas nas outras duas áreas parecem usar centros de inclusão digital mais frequentemente. Editar imagens e usar planilhas são atividades com proporção aparentemente maior nas interessadas em Exatas.

Figura 4 – Gráfico de mosaico para as variáveis Área de Interesse e Local de Uso de Computador (2019).



Fonte: Elaboração do autor. A: em casa; B: em casa de parentes; C: em casa de amigos; D: na escola; E: no trabalho; F: em *lan-house*; G: biblioteca; H: centros de inclusão digital; I: edição de texto; J: edição de imagem; K: planilha; L: banco de dados; M: acesso a *internet*; N: redes sociais; O: *e-mail*; P: jogos; Q: desenvolvimentos de páginas; R: programação e; S: outras atividades.

Tabela 16 – Tabela de contingência marginal para as variáveis Área de Interesse e Local de Uso de Computador (2019).

Local de Uso do Computador	Área de Interesse		
	Exatas	Biológicas e Saúde	Humanas
Em casa	47 (88,68%)	174 (82,86%)	111 (78,17%)
Em casa de parentes	27 (50,94%)	105 (50,00%)	62 (43,66%)
Em casa de amigos	16 (30,19%)	89 (42,38%)	44 (30,99%)
Na escola	14 (26,42%)	40 (19,05%)	35 (24,65%)
No trabalho	5 (9,43%)	24 (11,43%)	26 (18,31%)
Em <i>lan-house</i>	16 (30,19%)	52 (24,76%)	42 (29,58%)
Biblioteca	8 (15,09%)	24 (11,43%)	22 (15,49%)
Centros de inclusão digital	4 (7,55%)	6 (2,86%)	8 (5,63%)
Total de respondentes (m_{ij})	53 (100,00%)	210 (100,00%)	142 (100,00%)

Fonte: Elaboração do autor. Porcentagem das respostas em relação a m_{ij} em parênteses.

Tabela 17 – Tabela de contingência marginal para as variáveis Área de Interesse e Atividade Realizada no Computador (2019)

Atividade Realizada no Computador	Área de Interesse		
	Exatas	Biológicas e Saúde	Humanas
Edição de texto	41 (77,36%)	161 (76,67%)	111 (78,17%)
Edição de imagem	29 (54,72%)	97 (46,19%)	65 (45,77%)
Planilha	22 (41,51%)	75 (35,71%)	56 (39,44%)
Banco de dados	11 (20,75%)	35 (16,67%)	24 (16,90%)
Acesso a internet	50 (94,34%)	194 (92,38%)	132 (92,96%)
Redes sociais	47 (88,68%)	193 (91,90%)	132 (92,96%)
e-mail	44 (83,02%)	167 (79,52%)	117 (82,39%)
Jogos	38 (71,70%)	161 (76,67%)	109 (76,76%)
Desenvolvimento de páginas	5 (9,43%)	25 (11,90%)	23 (16,20%)
Programação	7 (13,21%)	28 (13,33%)	17 (11,97%)
Outros	12 (22,64%)	52 (24,76%)	29 (20,42%)
Total de respondentes (m_{ij})	53 (100,00%)	210 (100,00%)	142 (100,00%)

Fonte: Elaboração do autor. Porcentagem das respostas em relação a m_{ij} em parênteses.

Os resultados dos testes de MMI para as variáveis Área de Interesse vs. Local de Uso do Computador e Área de Interesse vs. Atividades Realizadas no Computador no período de 2019 relatam valores-p não significativos para dependência entre as variáveis. Não há dependência estatisticamente significativa da área de interesse das respondentes com os locais onde usam o computador ou com as atividades que realizam no computador. A saída R com os valores-p ajustados por Bonferroni para cada local e uso estão descritos nas Tabelas 18 e 19.

Tabela 18 – Saída R com as estatísticas Qui-Quadrado de Pearson não ajustadas e valores-p ajustados por Bonferroni para as variáveis Área de Interesse vs. Local de Uso do Computador (2019).

Local de Uso do Computador	Estatística Qui-Quadrado de Pearson não ajustada ($X_{S(ij)}^2$)	Valor-p ajustado por Bonferroni (p_{ij})
Em casa	3,11	1,0000
Em casa de parentes	1,59	1,0000
Em casa de amigos	5,87	0,4244
Na escola	2,25	1,0000
No trabalho	4,31	0,9266
Em <i>lan-house</i>	1,28	1,0000
Biblioteca	1,37	1,0000
Centros de inclusão digital	2,92	1,0000

Fonte: Elaboração do autor.

Tabela 19 – Saída R com as estatísticas Qui-Quadrado de Pearson não ajustadas e valores-p ajustados por Bonferroni para as variáveis Área de Interesse vs. Atividade Realizada no Computador (2019).

Atividade Realizada no Computador	Estatística Qui-Quadrado de Pearson não ajustada ($X_{S(ij)}^2$)	Valor-p ajustado por Bonferroni (p_{ij})
Edição de texto	0,11	1,0000
Edição de imagem	1,4	1,0000
Planilha	0,86	1,0000
Banco de dados	0,52	1,0000
Acesso a internet	0,25	1,0000
Redes sociais	0,95	1,0000
e-mail	0,62	1,0000
Jogos	0,63	1,0000
Desenvolvimento de páginas	2,09	1,0000
Programação	0,15	1,0000
Outros	0,91	1,0000

Fonte: Elaboração do autor.

Tabela 20 – Valores-p para os testes MMI e SPMI (2019).

Método	AdI		Local vs. Atividade			
	vs. Local	vs. Atividade	Geral	Exatas	Biol.S.	Humanas
Bonferroni	0,4244	1,0000	< 0,0001	0,0108	0,0017	0,0513
X_S^2 <i>bootstrap</i>	0,1336	0,991	< 0,0005	0,0925	< 0,0005	0,0325
Produto p_{ij} <i>bootstrap</i>	0,1336	0,991	< 0,0005	0,0765	< 0,0005	0,037
$\min_{ij} p_{ij}$ <i>bootstrap</i>	0,3387	0,988	0,0005	0,039	0,0065	0,073
X_{RS2}^2	0,1397	0,9849	-	-	-	-

Fonte: Elaboração do autor. Área de Interesse abreviado por AdI e Biológicas e Saúde abreviado por Biol.S.

Tabela 21 – Valores-p por ajuste de Bonferroni para cada estatística $X_{S(i,j)}^2$ para o teste SPMI das variáveis Local de Uso de Computador e Atividade Realizada no Computador (2019)

Local	Atividade										
	I	J	K	L	M	N	O	P	Q	R	S
A	1,000	0.114	1,000	1,000	1,000	1,000	0.075	1,000	1,000	1,000	1,000
B	1,000	1,000	1,000	1,000	1,000	1,000	1,000	0.004	1,000	1,000	1,000
C	1,000	0.756	1,000	1,000	1,000	0.630	1,000	0.034	1,000	1,000	1,000
D	1,000	0.338	1,000	1,000	1,000	1,000	1,000	1,000	1,000	1,000	1,000
E	0.089	1,000	0.000	0.099	1,000	1,000	0.156	1,000	1,000	1,000	1,000
F	1,000	1,000	1,000	1,000	1,000	1,000	1,000	0.108	1,000	1,000	1,000
G	1,000	1,000	1,000	1,000	1,000	1,000	1,000	1,000	1,000	1,000	0.075
H	1,000	1,000	1,000	1,000	1,000	1,000	1,000	1,000	1,000	1,000	1,000

Fonte: Elaboração do autor. A: em casa; B: em casa de parentes; C: em casa de amigos; D: na escola; E: no trabalho; F: em *lan-house*; G: biblioteca; H: centros de inclusão digital; I: edição de texto; J: edição de imagem; K: planilha; L: banco de dados; M: acesso a *internet*; N: redes sociais; O: *e-mail*; P: jogos; Q: desenvolvimentos de páginas; R: programação e; S: outras atividades.

Tabela 22 – Valores-p por ajuste de Bonferroni para cada estatística $X_{S(i,j)}^2$ para o teste SPMI das variáveis Local de Uso de Computador e Atividade Realizada no Computador para interessadas em Exatas (2011 - 2019)

Local	Atividade										
	I	J	K	L	M	N	O	P	Q	R	S
A	0,547	1,000	1,000	1,000	1,000	1,000	1,000	1,000	1,000	1,000	1,000
B	1,000	1,000	1,000	1,000	1,000	1,000	1,000	1,000	1,000	1,000	1,000
C	1,000	1,000	1,000	1,000	1,000	1,000	1,000	1,000	1,000	1,000	1,000
D	1,000	1,000	1,000	1,000	1,000	1,000	1,000	1,000	1,000	1,000	1,000
E	1,000	1,000	1,000	1,000	1,000	1,000	1,000	1,000	1,000	1,000	1,000
F	1,000	1,000	1,000	1,000	1,000	1,000	1,000	1,000	1,000	1,000	1,000
G	1,000	1,000	1,000	1,000	1,000	1,000	1,000	1,000	1,000	1,000	0,011
H	1,000	1,000	1,000	1,000	1,000	1,000	1,000	1,000	1,000	1,000	1,000

Fonte: Elaboração do autor. A: em casa; B: em casa de parentes; C: em casa de amigos; D: na escola; E: no trabalho; F: em *lan-house*; G: biblioteca; H: centros de inclusão digital; I: edição de texto; J: edição de imagem; K: planilha; L: banco de dados; M: acesso a *internet*; N: redes sociais; O: *e-mail*; P: jogos; Q: desenvolvimentos de páginas; R: programação e; S: outras atividades.

Tabela 23 – Valores-p por ajuste de Bonferroni para cada estatística $X_{S(i,j)}^2$ para o teste SPMI das variáveis Local de Uso de Computador e Atividade Realizada no Computador para interessadas em Biológicas e Saúde (2019)

Local	Atividade										
	I	J	K	L	M	N	O	P	Q	R	S
A	1,000	0,135	1,000	1,000	1,000	0,056	0,047	0,088	1,000	1,000	1,000
B	1,000	1,000	1,000	1,000	0,818	1,000	1,000	0,054	1,000	1,000	1,000
C	1,000	1,000	1,000	1,000	1,000	0,678	1,000	1,000	1,000	1,000	1,000
D	1,000	0,704	1,000	1,000	1,000	1,000	1,000	1,000	1,000	1,000	1,000
E	1,000	1,000	0,002	0,042	1,000	1,000	1,000	1,000	1,000	1,000	0,972
F	1,000	1,000	1,000	1,000	1,000	1,000	1,000	1,000	1,000	1,000	1,000
G	1,000	0,051	1,000	1,000	1,000	1,000	1,000	1,000	1,000	1,000	0,034
H	1,000	1,000	1,000	1,000	1,000	1,000	1,000	1,000	1,000	1,000	1,000

Fonte: Elaboração do autor. A: em casa; B: em casa de parentes; C: em casa de amigos; D: na escola; E: no trabalho; F: em *lan-house*; G: biblioteca; H: centros de inclusão digital; I: edição de texto; J: edição de imagem; K: planilha; L: banco de dados; M: acesso a *internet*; N: redes sociais; O: *e-mail*; P: jogos; Q: desenvolvimentos de páginas; R: programação e; S: outras atividades.

Tabela 24 – Valores-p por ajuste de Bonferroni para cada estatística $X_{S(i,j)}^2$ para o teste SPMI das variáveis Local de Uso de Computador e Atividade Realizada no Computador para interessadas em Humanas (2019)

Local	Atividade										
	I	J	K	L	M	N	O	P	Q	R	S
A	1,000	1,000	1,000	1,000	1,000	1,000	1,000	1,000	1,000	1,000	1,000
B	1,000	1,000	1,000	1,000	1,000	1,000	1,000	1,000	1,000	1,000	1,000
C	1,000	1,000	1,000	1,000	1,000	1,000	1,000	0,658	1,000	1,000	1,000
D	1,000	1,000	1,000	1,000	1,000	1,000	1,000	1,000	1,000	1,000	1,000
E	1,000	1,000	0,051	1,000	1,000	1,000	1,000	1,000	1,000	1,000	1,000
F	1,000	1,000	1,000	0,333	1,000	1,000	1,000	1,000	1,000	1,000	1,000
G	1,000	1,000	1,000	1,000	1,000	1,000	1,000	1,000	1,000	1,000	1,000
H	1,000	1,000	1,000	1,000	1,000	1,000	1,000	1,000	1,000	1,000	1,000

Fonte: Elaboração do autor. A: em casa; B: em casa de parentes; C: em casa de amigos; D: na escola; E: no trabalho; F: em *lan-house*; G: biblioteca; H: centros de inclusão digital; I: edição de texto; J: edição de imagem; K: planilha; L: banco de dados; M: acesso a *internet*; N: redes sociais; O: *e-mail*; P: jogos; Q: desenvolvimentos de páginas; R: programação e; S: outras atividades.

O teste SPMI conduzido para as variáveis Local de Uso do Computador vs. Atividades Realizadas no Computador indicam dependência estatisticamente significativa de ao menos uma atividade com um local de uso do computador para todos os cinco métodos. Há alguma dependência entre pelo menos uma combinação de local de uso de computador e atividade realizada no computador para as respondentes nesse período de 2019.

Cada célula das Tabelas 21 a 24 é o valor-p ajustado por Bonferroni para cada estatística $X_{S(i,j)}^2$ na condução do teste SPMI para as MRCVs Local de Uso do Computador e Atividade Realizada no Computador para o período de 2019 segundo respostas totais e respostas para respondentes interessadas nas áreas de Exatas, Biológicas e Saúde e

Humanas, respectivamente. Células para determinada atividade e determinado local com tonalidade de cinza mais escura indicam dependência estatisticamente significativa e células com fundo branco representam independência estatística entre os níveis das variáveis. Analisando a Tabela 21, jogar no computador em casa de parentes ou de amigos e usar planilhas no trabalho são combinações cuja dependência é significativa em geral. Pela Tabela 22, realizar outras atividades na biblioteca parece é uma combinação dependente entre as interessadas em Exatas. A partir da Tabela 23, usar planilha no trabalho é uma combinação dependente entre as interessadas em Biológicas e Saúde enquanto usar redes sociais e *e-mail* em casa, jogar jogos em casa de parentes, usar banco de dados no trabalho e editar imagens na biblioteca são combinações cuja dependência ou independência não se deve concluir por conta de seus valores-p. Segundo a Tabela 24, usar planilha no trabalho é a única combinação em que nada se deve concluir sobre a existência ou inexistência de dependência também por conta do valor-p. Um motivo admissível para o número expressivo de combinações cuja dependência ou independência não pode ser concluída para alguns níveis de Local de Uso do Computador e Atividade Realizada no Computador segundo diferentes níveis de Área de Interesse é a presença de muitas células com frequências baixíssimas.

3.3 Ajuste do modelo

Foram ajustados modelos log-lineares marginais generalizados para os dados coletados do questionário de Percepção sobre Computação do projeto Meninas da Computação. Para os ajustes destes modelos aos dados os 8 locais de uso do computador foram agrupados em quatro grupos (em casa; em casa de parentes ou amigos; na escola ou no trabalho e; em *lan-house*, na biblioteca ou em centros de inclusão digital) e as 11 atividades também foram agrupadas em quatro grupos (edição de texto ou imagem; planilha ou banco de dados; acesso a internet, redes sociais ou *e-mail* e; jogos, desenvolvimento de páginas, programação e outros). A junção de níveis para cada variável é necessária por conta da exigência alta do uso da memória do sistema ao utilizar a função do pacote MRCV que ajusta o modelo de interesse.

Os modelos considerados são Efeito Principal do Local, Efeito Principal da Atividade, Efeitos Principais do Local e da Atividade, SPMI e Homogêneo. O modelo Efeito Principal do Local ajusta o intercepto, os efeitos do Local e da Atividade e o efeito do Local dentro de cada nível da Atividade, permitindo associações entre os níveis do Local nas atividades. O modelo Efeito Principal da Atividade também ajusta o intercepto e os efeitos do Local e da Atividade, porém ajusta o efeito da Atividade dentro de cada nível do Local, permitindo associações entre os níveis da Atividade nos locais. O modelo Efeitos Principais do Local e da Atividade ajusta o intercepto, os efeitos do Local e da Atividade e ambos os efeitos do Local dentro de cada nível da Atividade e Atividade dentro de cada

nível do Local, permitindo associações entre os níveis do Local nas atividades e associações entre os níveis da Atividade nos locais. O modelo SPMI ajusta apenas o intercepto e os efeitos do Local e da Atividade. O modelo Homogêneo ajusta o intercepto, os efeitos do Local e da Atividade e um efeito de controle das margens das subtabelas 2×2 .

A Tabela 25 apresenta estatísticas de teste de máxima verossimilhança *bootstrap* para comparação de modelos log-lineares marginais generalizados para as variáveis Local de Uso do Computador e Atividades Realizadas no Computador. O uso de correção de segunda ordem de Rao-Scott retornou resultados conflituosos com a técnica *bootstrap*. Bilder e Loughin (2004) discutem a consistência e o poder de detecção de diversas hipóteses alternativas através dos métodos *bootstrap* bem como o fato do ajuste de Bonferroni e da correção de segunda ordem de Rao-Scott serem testes mais conservadores. Por esse motivo, apenas os resultados usando os métodos *bootstrap* são mostrados.

Tabela 25 – Estatísticas de teste de máxima verossimilhança *bootstrap* para comparação de modelos log-lineares marginais generalizados para as variáveis Local de Uso do Computador e Atividades Realizadas no Computador.

Modelo (H_0)	Modelo (H_A)					
	Loc.	Ativ.	Loc. e Ativ.	SPMI	Hmg.	Sat.
Loc.	-	1,60 (0,0505)	6,87 (0,039)	-49,71 (0,4117)	-4,97 (0,2971)	31,4 ($< 0,0005$)
Ativ.	-1,60 (0,2746)	-	5,26 (0,2586)	-51,31 (0,4442)	-6,57 (0,3647)	29,8 (0,0020)
Loc. e Ativ.	-6,87 (0,3582)	-5,26 (0,2826)	-	-56,57 (0,3652)	-11,83 (0,2121)	24,54 (0,0010)
SPMI	49,71 ($< 0,0005$)	51,31 ($< 0,0005$)	56,57 ($< 0,0005$)	-	44,74 ($< 0,0005$)	81,11 ($< 0,0005$)
Hmg.	4,97 (0,3082)	6,57 (0,0435)	11,83 (0,0900)	-44,74 (0,5038)	-	36,37 (0,0010)

Fonte: Elaboração do autor. Valores-p em parênteses. Local é abreviado por Loc., Atividade é abreviada por Ativ., Homogêneo é abreviado por Hmg. e Saturado é abreviado por Sat. Os graus de liberdade quando H_0 é o modelo X, Y, XY, SPMI e Homogêneo são 6,97, 7,69, 5,46, 9,58 e 9,2 respectivamente.

A comparação de modelos é feita usando o critério de informação de Akaike (AIC) cuja fórmula é dada por

$$AIC = G^2 - 2 \times df,$$

onde G^2 é a estatística de razão de verossimilhança e df são os graus de liberdade do modelo. O melhor modelo é aquele cujo AIC apresenta o menor valor. A Tabela 26 apresenta a estatística de razão de verossimilhança, os graus de liberdade e o AIC para cada modelo, comparando-os com o modelo saturado.

Sendo assim, podemos concluir que o modelo Efeito Principal da Atividade parece

ser o melhor dentre os candidatos e é descrito como

$$\log(\mu_{uv(ij)}) = \gamma_{ij} + \eta_u^X(ij) + \eta_v^Y(ij) + \lambda_{uv} + \lambda_{uv(j)}^Y,$$

onde γ_{ij} controla o total marginal da tabela, $\eta_u^X(ij)$ controla o total marginal de cada linha nas subtabelas 2×2 , $\eta_v^Y(ij)$ controla o total marginal de cada coluna nas subtabelas 2×2 e $\lambda_{uv(j)}^Y$ é o termo que permite variações entre os itens da variável Atividade Realizada no Computador. Isso indica que as atividades realizadas têm comportamentos distintos em locais de uso do computador.

Tabela 26 – Modelos ajustados, seus graus de liberdade, sua estatística de razão de verossimilhança e seu AIC comparado ao modelo saturado.

Modelo	G^2	Graus de liberdade	AIC
Local	31,40	12	7,40
Atividade	29,80	12	5,80
Local-Atividade	24,54	9	6,54
SPMI	81,11	16	49,11
Homogêneo	36,37	15	6,37

Fonte: Elaboração do autor.

4 Conclusão

O objetivo alvejado por este trabalho é introduzir a literatura acerca de variáveis categóricas de resposta múltipla (MRCV) e dados CATA e aplicar as abordagens encontradas para análise de problemas envolvendo esses tipos de variáveis e dados. Para alcançar este propósito, este trabalho discorreu sobre a terminologia ao longo dos anos, discutiu procedimentos como ajuste de Bonferroni, correções de segunda ordem de Rao-Scott e aproximações de *bootstrap* para testar independência marginal múltipla (MMI) e independência marginal em pares simultâneos (SPMI) e ajustes de modelos log-lineares generalizados para dados CATA e aplicou a metodologia revisada em um conjunto de dados real usando o pacote MRCV do *software* R.

Para as respostas de 2011 a 2014 do questionário Percepção sobre Computação do projeto Meninas na Computação, o teste de MMI indica dependência entre a área de interesse com os locais de uso do computador trabalho e *lan-house* e também dependência entre a área de interesse com as atividades edição de texto, uso de planilha, uso de banco de dados, jogar jogos no computador, desenvolvimento de páginas e programação. Como o valor-p ajustado por Bonferroni para as atividades acesso a *internet*, *e-mail* e outras atividades estão muito próximas do nível de confiança $\alpha = 5\%$ estipulado, não se pode concluir dependência ou independência entre essas atividades com a área de interesse das participantes do questionário. O teste SPMI para as variáveis Local de Uso do Computador e Atividade Realizada no Computador indica dependência entre muitos locais de uso e atividades realizadas para o período de 2011 a 2014.

Os testes de MMI para as respostas de 2019 do questionário indicam independência entre a área de interesse das respondentes com o local de uso do computador e atividades realizadas no computador. O teste SPMI ainda indica dependência entre local de uso e atividades realizadas. Jogar jogos de computador em casa de parentes ou de amigos e usar planilha no trabalho são combinações dependentes. Para interessadas em Exatas, realizar outras atividades na biblioteca é uma combinação dependente. Para interessadas em Biológicas e Saúde, usar a planilha no trabalho é uma combinação dependente.

Durante o uso do pacote MRCV, um obstáculo crucial foi encontrado. Não foi possível realizar os ajustes de modelos log-lineares generalizados das duas MRCVs com seus níveis originais (8 níveis da variável Local de Uso do Computador e 11 níveis da variável Atividades Realizadas no Computador) por conta do alto requerimento de memória do sistema. Neste caso, a junção de níveis em categorias com características semelhantes foi necessária para o prosseguimento do trabalho.

Dentre os modelos log-lineares ajustados, o modelo Y principal que ajusta o

intercepto e os efeitos das variáveis Local de Uso do Computador e Atividade Realizada no Computador e permite variações das associações entre os itens de Atividade Realizada no Computador, mantendo associações iguais entre os itens de Local de Uso do Computador é o melhor dentre os candidatos segundo o AIC. No entanto, os métodos com correção de segunda ordem de Rao-Scott mostraram resultados conflituosos com os métodos *bootstrap*. Preconizamos um estudo acerca de métodos que aplicam a correção de segunda ordem de Rao-Scott para verificar consistência e viabilidade.

O computador utilizado possui processador Intel(R) Core™ i3-3250 CPU @3.50GHz 3.50GHz, 8,00 GB de memória RAM instalada em Sistema Operacional Windows 10 Pro (2019) de 64 bits. O código de programa R utilizado na análise está disponível no Apêndice B, mas autorização para uso do banco de dados deve ser solicitado a Professora Maristela Terto de Holanda via *e-mail*¹.

Uma alternativa para o problema de memória do sistema é utilizar os programas SAS no *site* de Christopher R. Bilder². Os exemplos do *site* bem como os dados do questionário Percepção sobre Computação funcionam nos códigos SAS fornecidos por Bilder, porém os resultados das saídas são inúmeros. Sugerimos um estudo e organização da programação SAS com implementação de macros a fim de facilitar a interpretação de modelos log-lineares marginais generalizados com os níveis originais das MRCVs.

¹ Disponível em: <<http://www.meninas.cic.unb.br/contact>>. Acesso em: 27 nov. 2019.

² Disponível em: <<http://www.chrisbilder.com/mrcv/>>. Acesso em: 5 abr. 2019.

Referências

- AGRESTI, A.; LIU, I. Modeling a Categorical Variable Allowing Arbitrarily Many Category Choices. **Biometrics**, v. 55, p. 936–943, 1999.
- ATHREYA, K. B. Bootstrap of the Mean in the Infinite Variance Case. **The Annals of Statistics**, v. 15, n. 2, p. 724–731, 1987.
- BICKEL, P. J.; DOKSUM, K. A. **Mathematical Statistics: Basic Ideas and Selected Topics**. United States of America: Holden-Day, 1977.
- BILDER, C.; LOUGHIN, T. M.; NETTLETON, D. Multiple Marginal Independence Testing for Pick Any/C variables. **Communications in Statistics: Simulation and Computation**, v. 29, n. 4, p. 1285–1316, 2000.
- BILDER, C. R.; LOUGHIN, T. M. On the First-order Rao-Scott Correction of the Umesh-Loughin-Scherer Statistic. **Biometrics**, v. 57, p. 1253–1255, 2001.
- BILDER, C. R.; LOUGHIN, T. M. Testing for Conditional Multiple Marginal Independence. **Biometrics**, v. 58, p. 200–208, 2002a.
- BILDER, C. R.; LOUGHIN, T. M. Testing for Simultaneous Pairwise Multiple Marginal Independence. **Joint Statistical Meetings - Biometrics Section-to include ENAR & WNAR**, p. 254–259, 2002b.
- BILDER, C. R.; LOUGHIN, T. M. Strategies for Modeling Two Categorical Variables with Multiple Category Choices. **Joint Statistical Meetings - Biometrics Section-to include ENAR & WNAR**, p. 560–567, 2003.
- BILDER, C. R.; LOUGHIN, T. M. Testing for Marginal Independence between Two Categorical Variables with Multiple Responses. **Biometrics**, v. 60, p. 241–248, 2004.
- BILDER, C. R.; LOUGHIN, T. M. Modeling multiple-response categorical data from complex surveys. **The Canadian Journal of Statistics**, v. 37, n. 4, p. 553–570, 2009.
- CABIN, R. J.; MITCHELL, R. J. To Bonferroni or Not to Bonferroni: When and How Are the Questions. **ESA Bull**, v. 81, p. 246–248, 2000.
- CLAESKENS, G.; HJORT, N. **Model Selection And Model Averaging**. Cambridge: Cambridge University Press, 2008.
- DECADY, Y. J.; THOMAS, D. H. A Simple Test of Association for Contingency Tables with Multiple Column Responses. **Biometrics**, v. 56, p. 893–896, 2000.
- EFRON, B. Computers and the Theory of Statistics: Thinking the Unthinkable. **Society for Industrial and Applied Mathematics**, v. 21, n. 4, p. 460–480, 10 1979.
- EFRON, B. **The Jackknife, the Bootstrap and Other Resampling Plans**. 1. ed. United States of America: Society for Industrial and Applied Mathematics, 1982.
- EFRON, B.; TIBSHIRANI, R. **An Introduction to the Bootstrap**. 1. ed. [S.l.]: Chapman & Hall/CRC, 1999.

- HOLANDA, M. T.; ARAUJO, A. P. F. de; WALTER, M. E. M. T. **Meninas na Computação**. 2011 – atual. Projeto de Pesquisa conduzido pelo Departamento de Computação da Universidade de Brasília - meninas.cic.unb.br.
- KOZIOL, N. A.; BILDER, C. R. MRCV: A Package for Analyzing Categorical Variables with Multiple Response Options. **The R Journal**, v. 6, n. 1, p. 144–150, 2014.
- LEVINE, J. H. Joint-Space Analysis of “Pick-Any” Data: Analysis of Choices from an Unconstrained Set of Alternatives. **Psychometrika**, The Psychometric Society, v. 44, n. 1, p. 85–92, 3 1979.
- LOHR, S. L. **Sampling: Design and Analysis**. 2. ed. Boston: Brooks/Cole Cengage Learning, 2009.
- LOUGHIN, T.; SCHERER, P. Testing for Association in Contingency Tables with Multiple Column Responses. **Biometrics**, Wiley, v. 54, n. 2, p. 630–637, 6 1998.
- NAKAGAWA, S. A farewell to Bonferroni: the problems of low statistical power and publication bias. **Behavioral Ecology**, v. 15, n. 6, p. 1044–1045, 11 2004.
- RAO, J. N. K.; SCOTT, A. J. The Analysis of Categorical Data From Complex Sample Surveys: Chi-Squared Tests for Goodness of Fit and Independence in Two-Way Tables. **Journal of the American Statistical Association**, v. 76, n. 374, p. 221–230, 6 1981.
- RAO, J. N. K.; SCOTT, A. J. On Chi-Squared Tests for Multiway Contingency Tables with Cell Proportions Estimated from Survey Data. **Institute of Mathematical Statistics**, v. 12, n. 1, p. 46–60, 3 1984.
- SATTERTHWAITE, F. E. An Approximate Distribution of Estimates of Variance Components. **Biometrics Bulletin**, v. 2, n. 6, p. 110–114, 12 1946.
- UMESH, U. N. Predicting nominal variable relationships with multiple response. **Journal of Forecasting**, v. 14, p. 585–596, 1995.
- WESTFALL, P. H.; YOUNG, S. S. **Resampling-Based Multiple Testing: Examples and Methods for p-Value Adjustment**. United States of America: Wiley, 1993.

Apêndices

APÊNDICE A – Tabelas de Resposta ao Item

Tabela 27 – Tabela de resposta ao item para as variáveis Área de Interesse e Local de Uso de Computador (2011 - 2014).

Área de Interesse	Local de uso do computador															
	A		B		C		D		E		F		G		H	
	1	0	1	0	1	0	1	0	1	0	1	0	1	0	1	0
Exat.	754	47	536	265	525	276	399	402	151	650	256	545	215	586	58	743
Bio. & S.	1361	118	1009	470	948	531	666	813	235	1244	553	926	365	1114	91	1388
Hum.	1164	86	809	441	823	427	585	665	322	928	377	873	307	943	93	1157

Fonte: Elaboração do autor. A: em casa; B: em casa de parentes; C: em casa de amigos; D: na escola; E: no trabalho; F: em *lan-house*; G: biblioteca e; H: centros de inclusão digital.

Tabela 28 – Tabela de resposta ao item para as variáveis Área de Interesse e Local de Uso de Computador (2019).

Área de Interesse	Local de uso do computador															
	A		B		C		D		E		F		G		H	
	1	0	1	0	1	0	1	0	1	0	1	0	1	0	1	0
Exat.	47	6	27	26	16	37	14	39	5	48	16	37	8	45	4	49
Bio. & S.	174	36	105	105	89	121	40	170	24	186	52	158	24	186	6	204
Hum.	111	31	62	80	44	98	35	107	26	116	42	100	22	120	8	134

Fonte: Elaboração do autor. A: em casa; B: em casa de parentes; C: em casa de amigos; D: na escola; E: no trabalho; F: em *lan-house*; G: biblioteca e; H: centros de inclusão digital.

Tabela 29 – Tabela de resposta ao item para as variáveis Área de Interesse e Atividade Realizada no Computador (2011 - 2014).

Área de Interesse	Atividades realizadas no computador																					
	I	J	K	L	M	N	O	P	Q	R	S											
Exat.	716	85	607	194	413	388	207	594	736	65	740	61	710	91	586	215	169	632	230	571	261	540
Bio. & S.	1239	240	1038	441	605	874	258	1221	1306	173	1352	127	1259	220	1000	479	190	1289	281	1198	388	1091
Hum.	1090	160	927	323	578	672	265	985	1144	106	1145	105	1114	136	801	449	180	1070	262	988	370	880

Fonte: Elaboração do autor. I: edição de texto; J: edição de imagem; K: planilha; L: banco de dados; M: acesso a *internet*; N: redes sociais; O: *e-mail*; P: jogos; Q: desenvolvimentos de páginas; R: programação e; S: outras atividades.

Tabela 30 – Tabela de resposta ao item para as variáveis Área de Interesse e Atividade Realizada no Computador (2019).

Área de Interesse	Atividades realizadas no computador																					
	I	J	K	L	M	N	O	P	Q	R	S											
Exat.	41	12	29	24	22	31	11	42	50	3	47	6	44	9	38	15	5	48	7	46	12	41
Bio. & S.	161	49	97	113	75	135	35	175	194	16	193	17	167	43	161	49	25	185	28	182	52	158
Hum.	111	31	65	77	56	86	24	118	132	10	132	10	117	25	109	33	23	119	17	125	29	113

Fonte: Elaboração do autor. I: edição de texto; J: edição de imagem; K: planilha; L: banco de dados; M: acesso a *internet*; N: redes sociais; O: *e-mail*; P: jogos; Q: desenvolvimentos de páginas; R: programação e; S: outras atividades.

Tabela 31 – Tabela de contingência marginal para as variáveis Local de Uso de Computador e Atividade Realizada no Computador (2011 - 2014).

Local	Atividade										
	I	J	K	L	M	N	O	P	Q	R	S
A	2892 (0,819)	2461 (0,697)	1514 (0,429)	682 (0,193)	2985 (0,846)	3043 (0,862)	2919 (0,827)	2234 (0,633)	509 (0,144)	739 (0,209)	959 (0,272)
B	2091 (0,592)	1850 (0,524)	1065 (0,302)	522 (0,148)	2163 (0,613)	2228 (0,631)	2125 (0,602)	1742 (0,494)	393 (0,111)	566 (0,160)	754 (0,214)
C	2055 (0,582)	1816 (0,514)	1068 (0,303)	509 (0,144)	2127 (0,603)	2197 (0,622)	2088 (0,592)	1690 (0,479)	396 (0,112)	560 (0,159)	748 (0,212)
D	1513 (0,429)	1312 (0,372)	857 (0,243)	404 (0,114)	1523 (0,431)	1548 (0,439)	1532 (0,434)	1189 (0,337)	314 (0,089)	420 (0,119)	556 (0,158)
E	667 (0,189)	562 (0,159)	520 (0,147)	261 (0,074)	667 (0,189)	660 (0,187)	675 (0,191)	457 (0,130)	150 (0,043)	174 (0,049)	242 (0,069)
F	1009 (0,286)	854 (0,242)	529 (0,150)	286 (0,081)	1063 (0,301)	1094 (0,310)	1013 (0,287)	904 (0,256)	211 (0,060)	276 (0,078)	411 (0,116)
G	823 (0,233)	736 (0,209)	504 (0,143)	264 (0,075)	825 (0,234)	826 (0,234)	829 (0,235)	664 (0,188)	190 (0,054)	268 (0,076)	334 (0,095)
H	225 (0,064)	188 (0,053)	152 (0,043)	105 (0,030)	220 (0,062)	226 (0,064)	218 (0,062)	187 (0,053)	78 (0,022)	110 (0,031)	123 (0,035)

Fonte: Elaboração do autor. Valores em parênteses são as proporções da atividade realizada no computador em cada área de interesse. O valor de referência para as proporções é $m_{ij} = 3530$. A: em casa; B: em casa de parentes; C: em casa de amigos; D: na escola; E: no trabalho; F: em *lan-house*; G: biblioteca; H: centros de inclusão digital; I: edição de texto; J: edição de imagem; K: planilha; L: banco de dados; M: acesso a *internet*; N: redes sociais; O: *e-mail*; P: jogos; Q: desenvolvimentos de páginas; R: programação e; S: outras atividades.

Tabela 32 – Tabela de resposta ao item para as variáveis Local de Uso de Computador e Atividade Realizada no Computador (2011 - 2014).

		Atividades realizadas no computador																					
		I	J	K	L	M	N	O	P	Q	R	S											
A	1	2892	387	2461	818	1514	1765	682	2597	2985	294	3043	236	2919	360	2234	1045	509	2770	739	2540	959	2320
	0	153	98	111	140	82	169	48	203	201	50	194	57	164	87	153	98	30	221	34	217	60	191
B	1	2091	263	1850	504	1065	1289	522	1832	2163	191	2228	126	2125	229	1742	612	393	1961	566	1788	754	1600
	0	954	222	722	454	531	645	208	968	1023	153	1009	167	958	218	645	531	146	1030	207	969	265	911
C	1	2055	241	1816	480	1068	1228	509	1787	2127	169	2197	99	2088	208	1690	606	396	1900	560	1736	748	1548
	0	990	244	756	478	528	706	221	1013	1059	175	1040	194	995	239	697	537	143	1091	213	1021	271	963
D	1	1513	137	1312	338	857	793	404	1246	1523	127	1548	102	1532	118	1189	461	314	1336	420	1230	556	1094
	0	1532	348	1260	620	739	1141	326	1554	1663	217	1689	191	1551	329	1198	682	225	1655	353	1527	463	1417
E	1	667	41	562	146	520	188	261	447	667	41	660	48	675	33	457	251	150	558	174	534	242	466
	0	2378	444	2010	812	1076	1746	469	2353	2519	303	2577	245	2408	414	1930	892	389	2433	599	2223	777	2045
F	1	1009	177	854	332	529	657	286	900	1063	123	1094	92	1013	173	904	282	211	975	276	910	411	775
	0	2036	308	1718	626	1067	1277	444	1900	2123	221	2143	201	2070	274	1483	861	328	2016	497	1847	608	1736
G	1	823	64	736	151	504	383	264	623	825	62	826	61	829	58	664	223	190	697	268	619	334	553
	0	2222	421	1836	807	1092	1551	466	2177	2361	282	2411	232	2254	389	1723	920	349	2294	505	2138	685	1958
H	1	225	17	188	54	152	90	105	137	220	22	226	16	218	24	187	55	78	164	110	132	123	119
	0	2820	468	2384	904	1444	1844	625	2663	2966	322	3011	277	2865	423	2200	1088	461	2827	663	2625	896	2392

Fonte: Elaboração do autor. A: em casa; B: em casa de parentes; C: em casa de amigos; D: na escola; E: no trabalho; F: em *lan-house*; G: biblioteca; H: centros de inclusão digital; I: edição de texto; J: edição de imagem; K: planilha; L: banco de dados; M: acesso a *internet*; N: redes sociais; O: *e-mail*; P: jogos; Q: desenvolvimento de páginas; R: programação e; S: outras atividades.

Tabela 33 – Tabela de contingência marginal para as variáveis Local de Uso de Computador e Atividade Realizada no Computador para interessadas em Exatas (2011 - 2014).

Local	Atividade										
	I	J	K	L	M	N	O	P	Q	R	S
A	680 (0,849)	588 (0,734)	395 (0,493)	198 (0,247)	697 (0,870)	703 (0,878)	685 (0,855)	561 (0,700)	164 (0,205)	224 (0,280)	247 (0,308)
B	494 (0,617)	435 (0,543)	279 (0,348)	140 (0,175)	500 (0,624)	513 (0,640)	497 (0,621)	417 (0,521)	120 (0,150)	159 (0,199)	195 (0,243)
C	483 (0,603)	431 (0,538)	280 (0,350)	135 (0,169)	487 (0,608)	501 (0,626)	483 (0,603)	412 (0,514)	124 (0,155)	191 (0,239)	748 (0,212)
D	372 (0,464)	329 (0,411)	243 (0,303)	132 (0,165)	370 (0,462)	378 (0,472)	373 (0,466)	308 (0,385)	99 (0,124)	137 (0,171)	150 (0,187)
E	149 (0,186)	122 (0,152)	119 (0,149)	71 (0,089)	144 (0,180)	139 (0,174)	147 (0,184)	111 (0,139)	41 (0,051)	51 (0,064)	63 (0,079)
F	225 (0,281)	188 (0,235)	123 (0,154)	71 (0,089)	232 (0,290)	243 (0,303)	226 (0,282)	210 (0,262)	59 (0,074)	72 (0,090)	100 (0,125)
G	209 (0,261)	192 (0,240)	154 (0,192)	87 (0,109)	201 (0,251)	202 (0,252)	208 (0,260)	174 (0,217)	62 (0,077)	97 (0,121)	89 (0,111)
H	56 (0,070)	50 (0,062)	40 (0,050)	35 (0,044)	56 (0,070)	57 (0,071)	52 (0,065)	49 (0,061)	25 (0,031)	31 (0,039)	31 (0,039)

Fonte: Elaboração do autor. Valores em parênteses são as proporções da atividade realizada no computador em cada área de interesse. O valor de referência para as proporções é $m_{ij} = 801$. A: em casa; B: em casa de parentes; C: em casa de amigos; D: na escola; E: no trabalho; F: em *lan-house*; G: biblioteca; H: centros de inclusão digital; I: edição de texto; J: edição de imagem; K: planilha; L: banco de dados; M: acesso a *internet*; N: redes sociais; O: *e-mail*; P: jogos; Q: desenvolvimentos de páginas; R: programação e; S: outras atividades.

Tabela 34 – Tabela de contingência marginal para as variáveis Local de Uso de Computador e Atividade Realizada no Computador para interessadas em Biológicas e Saúde (2011 - 2014).

Local	Atividade										
	I	J	K	L	M	N	O	P	Q	R	S
A	1177 (0,796)	985 (0,667)	572 (0,387)	240 (0,162)	1219 (0,824)	1264 (0,855)	1185 (0,801)	923 (0,624)	176 (0,119)	265 (0,179)	361 (0,244)
B	873 (0,590)	767 (0,519)	417 (0,282)	188 (0,127)	908 (0,614)	943 (0,638)	887 (0,600)	746 (0,504)	144 (0,097)	213 (0,144)	286 (0,193)
C	836 (0,565)	730 (0,494)	399 (0,270)	185 (0,125)	862 (0,583)	903 (0,611)	853 (0,577)	693 (0,469)	141 (0,095)	204 (0,138)	284 (0,192)
D	603 (0,408)	511 (0,346)	311 (0,210)	127 (0,086)	606 (0,410)	620 (0,419)	608 (0,411)	495 (0,385)	107 (0,072)	138 (0,093)	215 (0,145)
E	216 (0,146)	183 (0,124)	162 (0,110)	78 (0,053)	219 (0,148)	221 (0,149)	220 (0,149)	157 (0,106)	46 (0,031)	56 (0,038)	77 (0,052)
F	463 (0,313)	388 (0,262)	224 (0,152)	116 (0,078)	489 (0,331)	503 (0,340)	458 (0,310)	419 (0,283)	86 (0,058)	109 (0,074)	171 (0,116)
G	331 (0,224)	293 (0,198)	173 (0,117)	89 (0,060)	335 (0,227)	336 (0,227)	333 (0,225)	270 (0,183)	68 (0,046)	89 (0,060)	138 (0,093)
H	84 (0,057)	67 (0,045)	55 (0,037)	34 (0,023)	80 (0,054)	85 (0,058)	82 (0,055)	71 (0,048)	28 (0,019)	35 (0,024)	48 (0,033)

Fonte: Elaboração do autor. Valores em parênteses são as proporções da atividade realizada no computador em cada área de interesse. O valor de referência para as proporções é $m_{ij} = 1479$. A: em casa; B: em casa de parentes; C: em casa de amigos; D: na escola; E: no trabalho; F: em *lan-house*; G: biblioteca; H: centros de inclusão digital; I: edição de texto; J: edição de imagem; K: planilha; L: banco de dados; M: acesso a *internet*; N: redes sociais; O: *e-mail*; P: jogos; Q: desenvolvimentos de páginas; R: programação e; S: outras atividades.

Tabela 35 – Tabela de contingência marginal para as variáveis Local de Uso de Computador e Atividade Realizada no Computador para interessadas em Humanas (2011 - 2014).

Local	Atividade										
	I	J	K	L	M	N	O	P	Q	R	S
A	1035 (0,828)	888 (0,710)	547 (0,438)	244 (0,195)	1069 (0,855)	1076 (0,861)	1049 (0,839)	750 (0,600)	169 (0,135)	250 (0,200)	351 (0,281)
B	724 (0,579)	648 (0,518)	369 (0,295)	194 (0,155)	755 (0,604)	772 (0,618)	741 (0,593)	579 (0,463)	129 (0,103)	194 (0,155)	273 (0,218)
C	736 (0,589)	655 (0,524)	389 (0,311)	189 (0,151)	778 (0,622)	793 (0,634)	752 (0,602)	585 (0,468)	131 (0,105)	192 (0,154)	273 (0,218)
D	538 (0,430)	472 (0,377)	303 (0,242)	145 (0,116)	547 (0,438)	550 (0,440)	551 (0,441)	386 (0,309)	108 (0,086)	145 (0,116)	191 (0,153)
E	302 (0,242)	257 (0,206)	239 (0,191)	112 (0,090)	304 (0,243)	300 (0,240)	308 (0,246)	189 (0,151)	63 (0,050)	67 (0,054)	102 (0,082)
F	321 (0,257)	278 (0,222)	182 (0,146)	99 (0,079)	342 (0,274)	348 (0,278)	329 (0,263)	275 (0,220)	66 (0,053)	95 (0,076)	140 (0,112)
G	283 (0,226)	251 (0,201)	177 (0,142)	88 (0,070)	289 (0,231)	288 (0,230)	288 (0,230)	220 (0,176)	60 (0,048)	82 (0,066)	107 (0,086)
H	85 (0,068)	71 (0,057)	57 (0,046)	36 (0,029)	84 (0,067)	84 (0,067)	84 (0,067)	67 (0,054)	25 (0,020)	44 (0,035)	44 (0,035)

Fonte: Elaboração do autor. Valores em parênteses são as proporções da atividade realizada no computador em cada área de interesse. O valor de referência para as proporções é $m_{ij} = 1250$. A: em casa; B: em casa de parentes; C: em casa de amigos; D: na escola; E: no trabalho; F: em *lan-house*; G: biblioteca; H: centros de inclusão digital; I: edição de texto; J: edição de imagem; K: planilha; L: banco de dados; M: acesso a *internet*; N: redes sociais; O: *e-mail*; P: jogos; Q: desenvolvimentos de páginas; R: programação e; S: outras atividades.

Tabela 36 – Tabela de contingência marginal para as variáveis Locais de Uso de Computador e Atividade Realizada no Computador (2019).

Local	Atividade										
	I	J	K	L	M	N	O	P	Q	R	S
A	263 (0,649)	169 (0,417)	129 (0,319)	63 (0,156)	312 (0,770)	310 (0,765)	279 (0,689)	258 (0,637)	47 (0,116)	46 (0,114)	84 (0,207)
B	154 (0,380)	99 (0,244)	76 (0,188)	34 (0,084)	186 (0,459)	182 (0,449)	163 (0,403)	165 (0,407)	30 (0,074)	25 (0,062)	52 (0,128)
C	122 (0,301)	83 (0,205)	55 (0,136)	29 (0,072)	143 (0,353)	144 (0,356)	125 (0,309)	128 (0,316)	24 (0,059)	26 (0,064)	38 (0,094)
D	75 (0,185)	54 (0,133)	36 (0,089)	19 (0,047)	84 (0,207)	85 (0,210)	80 (0,198)	73 (0,180)	12 (0,030)	13 (0,032)	28 (0,069)
E	52 (0,128)	31 (0,077)	40 (0,099)	18 (0,044)	53 (0,131)	51 (0,126)	53 (0,131)	36 (0,089)	11 (0,027)	9 (0,022)	14 (0,035)
F	84 (0,207)	54 (0,133)	37 (0,091)	22 (0,054)	105 (0,259)	102 (0,252)	92 (0,227)	96 (0,237)	15 (0,037)	16 (0,040)	22 (0,054)
G	46 (0,207)	34 (0,133)	24 (0,091)	11 (0,054)	51 (0,259)	46 (0,252)	46 (0,227)	43 (0,237)	9 (0,037)	9 (0,040)	22 (0,054)
H	13 (0,032)	11 (0,027)	11 (0,027)	5 (0,012)	18 (0,044)	17 (0,042)	16 (0,040)	14 (0,035)	3 (0,007)	4 (0,010)	6 (0,015)

Fonte: Elaboração do autor. Valores em parênteses são as proporções da atividade realizada no computador em cada área de interesse. O valor de referência para as proporções é $m_{ij} = 405$. A: em casa; B: em casa de parentes; C: em casa de amigos; D: na escola; E: no trabalho; F: em *lan-house*; G: biblioteca; H: centros de inclusão digital; I: edição de texto; J: edição de imagem; K: planilha; L: banco de dados; M: acesso a *internet*; N: redes sociais; O: *e-mail*; P: jogos; Q: desenvolvimentos de páginas; R: programação e; S: outras atividades.

Tabela 37 – Tabela de resposta ao item para as variáveis Locais de Uso de Computador e Atividade Realizada no Computador (2019).

		Atividades realizadas no computador																					
		I	J	K	L	M	N	O	P	Q	R	S											
A	1	0	1	0	1	0	1	0	1	0	1	0	1	0	1	0	1	0	1	0	1	0	
	0	263	69	169	163	129	203	63	269	312	20	310	22	279	53	258	74	47	285	46	286	84	248
B	1	0	23	22	51	24	49	7	66	64	9	62	11	49	24	50	23	6	67	6	67	9	64
	0	154	40	99	95	76	118	34	160	186	8	182	12	163	31	165	29	30	164	25	169	52	142
C	1	0	52	92	119	77	134	36	175	190	21	190	21	165	46	143	68	23	188	27	184	41	170
	0	122	27	83	66	55	94	29	120	143	6	144	5	125	24	128	21	24	125	26	123	38	111
D	1	0	65	108	148	98	158	41	215	233	23	228	28	203	53	180	76	29	227	26	230	55	201
	0	75	14	54	35	36	53	19	70	84	5	85	4	80	9	73	16	12	77	13	76	28	61
E	1	0	78	137	179	117	199	51	265	292	24	287	29	248	68	235	81	41	275	39	277	65	251
	0	52	3	31	24	40	15	18	37	53	2	51	4	53	2	36	19	11	44	9	46	14	41
F	1	0	89	160	190	113	237	52	298	323	27	321	29	275	75	272	78	42	308	43	307	79	271
	0	84	26	54	56	37	73	22	88	105	5	102	8	92	18	96	14	15	95	16	94	22	88
G	1	0	66	137	158	116	179	48	247	271	24	270	25	236	59	212	83	38	257	36	259	71	224
	0	46	8	34	20	24	30	11	43	51	3	46	8	46	8	43	11	9	45	9	45	22	32
H	1	0	84	157	194	129	222	59	292	325	26	326	25	282	69	265	86	44	307	43	308	71	280
	0	13	5	11	7	11	7	5	13	18	0	17	1	16	2	14	4	3	15	4	14	6	12
	0	300	87	180	207	142	245	65	322	358	29	355	32	312	75	294	93	50	337	48	339	87	300

Fonte: Elaboração do autor. A: em casa; B: em casa de parentes; C: em casa de amigos; D: na escola; E: no trabalho; F: em *lan-house*; G: biblioteca; H: centros de inclusão digital; I: edição de texto; J: edição de imagem; K: planilha; L: banco de dados; M: acesso a *internet*; N: redes sociais; O: *e-mail*; P: jogos; Q: desenvolvimento de páginas; R: programação e; S: outras atividades.

Tabela 38 – Tabela de contingência marginal para as variáveis Local de Uso de Computador e Atividade Realizada no Computador para a área de interesse em Exatas (2019).

Local	Atividade										
	I	J	K	L	M	N	O	P	Q	R	S
A	39 (0,736)	27 (0,509)	20 (0,377)	11 (0,208)	45 (0,849)	42 (0,793)	40 (0,755)	33 (0,623)	5 (0,094)	6 (0,113)	12 (0,226)
B	21 (0,396)	17 (0,321)	13 (0,245)	7 (0,132)	26 (0,491)	23 (0,434)	24 (0,453)	22 (0,415)	3 (0,057)	3 (0,057)	8 (0,151)
C	14 (0,264)	11 (0,208)	8 (0,151)	5 (0,094)	16 (0,302)	15 (0,283)	14 (0,264)	15 (0,283)	2 (0,038)	2 (0,038)	5 (0,094)
D	13 (0,245)	9 (0,170)	7 (0,132)	2 (0,038)	14 (0,264)	13 (0,245)	13 (0,245)	13 (0,245)	0 (0,000)	0 (0,000)	4 (0,076)
E	5 (0,094)	4 (0,076)	4 (0,076)	2 (0,038)	5 (0,094)	5 (0,094)	5 (0,094)	4 (0,076)	0 (0,000)	0 (0,000)	2 (0,038)
F	10 (0,199)	7 (0,132)	4 (0,076)	3 (0,057)	15 (0,283)	13 (0,245)	13 (0,245)	14 (0,264)	1 (0,019)	1 (0,019)	0 (0,000)
G	6 (0,113)	5 (0,094)	5 (0,094)	1 (0,019)	7 (0,013)	7 (0,013)	6 (0,113)	7 (0,132)	0 (0,000)	0 (0,000)	6 (0,113)
H	3 (0,057)	3 (0,057)	3 (0,057)	1 (0,019)	4 (0,076)	4 (0,076)	3 (0,057)	3 (0,057)	0 (0,000)	0 (0,000)	1 (0,019)

Fonte: Elaboração do autor. Valores em parênteses são as proporções da atividade realizada no computador em cada área de interesse. O valor de referência para as proporções é $m_{ij} = 53$. A: em casa; B: em casa de parentes; C: em casa de amigos; D: na escola; E: no trabalho; F: em *lan-house*; G: biblioteca; H: centros de inclusão digital; I: edição de texto; J: edição de imagem; K: planilha; L: banco de dados; M: acesso a *internet*; N: redes sociais; O: *e-mail*; P: jogos; Q: desenvolvimentos de páginas; R: programação e; S: outras atividades.

Tabela 39 – Tabela de contingência marginal para as variáveis Local de Uso de Computador e Atividade Realizada no Computador para a área de interesse em Biológicas e Saúde (2019).

Local	Atividade										
	I	J	K	L	M	N	O	P	Q	R	S
A	136 (0,648)	89 (0,424)	65 (0,310)	34 (0,162)	163 (0,776)	165 (0,786)	146 (0,695)	141 (0,671)	23 (0,110)	26 (0,124)	47 (0,224)
B	82 (0,391)	52 (0,248)	37 (0,176)	14 (0,067)	102 (0,486)	100 (0,476)	87 (0,414)	91 (0,433)	16 (0,076)	16 (0,076)	32 (0,152)
C	72 (0,343)	46 (0,219)	31 (0,148)	12 (0,057)	84 (0,400)	87 (0,414)	72 (0,343)	73 (0,348)	13 (0,062)	15 (0,071)	23 (0,110)
D	34 (0,162)	26 (0,124)	15 (0,071)	10 (0,048)	37 (0,176)	39 (0,186)	36 (0,171)	32 (0,152)	7 (0,033)	14 (0,067)	215 (0,145)
E	22 (0,105)	14 (0,067)	18 (0,086)	10 (0,048)	22 (0,105)	22 (0,105)	23 (0,110)	15 (0,071)	5 (0,024)	5 (0,024)	11 (0,052)
F	40 (0,191)	23 (0,110)	18 (0,086)	6 (0,029)	50 (0,238)	50 (0,238)	41 (0,195)	44 (0,210)	5 (0,024)	7 (0,033)	11 (0,052)
G	20 (0,095)	19 (0,091)	10 (0,048)	5 (0,024)	23 (0,110)	20 (0,095)	21 (0,100)	18 (0,086)	4 (0,019)	5 (0,024)	13 (0,062)
H	5 (0,024)	4 (0,019)	4 (0,019)	2 (0,010)	6 (0,029)	5 (0,024)	6 (0,029)	5 (0,024)	2 (0,010)	2 (0,010)	4 (0,019)

Fonte: Elaboração do autor. Valores em parênteses são as proporções da atividade realizada no computador em cada área de interesse. O valor de referência para as proporções é $m_{ij} = 210$. A: em casa; B: em casa de parentes; C: em casa de amigos; D: na escola; E: no trabalho; F: em *lan-house*; G: biblioteca; H: centros de inclusão digital; I: edição de texto; J: edição de imagem; K: planilha; L: banco de dados; M: acesso a *internet*; N: redes sociais; O: *e-mail*; P: jogos; Q: desenvolvimentos de páginas; R: programação e; S: outras atividades.

Tabela 40 – Tabela de contingência marginal para as variáveis Local de Uso de Computador e Atividade Realizada no Computador para a área de interesse em Humanas (2019).

Local	Atividade										
	I	J	K	L	M	N	O	P	Q	R	S
A	88 (0,620)	53 (0,373)	44 (0,310)	18 (0,127)	104 (0,732)	103 (0,725)	93 (0,655)	84 (0,592)	19 (0,134)	14 (0,099)	25 (0,176)
B	51 (0,359)	30 (0,211)	26 (0,183)	13 (0,092)	58 (0,409)	59 (0,416)	52 (0,366)	52 (0,366)	11 (0,078)	6 (0,042)	12 (0,085)
C	36 (0,254)	26 (0,183)	16 (0,113)	12 (0,085)	43 (0,303)	42 (0,296)	39 (0,275)	40 (0,282)	9 (0,063)	9 (0,063)	10 (0,070)
D	28 (0,197)	19 (0,134)	14 (0,099)	7 (0,049)	33 (0,232)	33 (0,232)	31 (0,218)	28 (0,197)	5 (0,035)	6 (0,042)	10 (0,070)
E	25 (0,176)	13 (0,092)	18 (0,127)	6 (0,042)	26 (0,183)	24 (0,169)	25 (0,176)	17 (0,120)	6 (0,042)	4 (0,028)	1 (0,007)
F	34 (0,239)	24 (0,169)	15 (0,106)	13 (0,092)	40 (0,282)	39 (0,275)	38 (0,268)	38 (0,268)	9 (0,063)	8 (0,056)	11 (0,078)
G	20 (0,141)	10 (0,070)	9 (0,063)	5 (0,035)	21 (0,148)	19 (0,134)	19 (0,134)	18 (0,127)	5 (0,035)	4 (0,028)	3 (0,021)
H	5 (0,035)	4 (0,028)	4 (0,028)	2 (0,014)	8 (0,056)	8 (0,056)	7 (0,049)	6 (0,042)	1 (0,007)	2 (0,014)	1 (0,007)

Fonte: Elaboração do autor. Valores em parênteses são as proporções da atividade realizada no computador em cada área de interesse. O valor de referência para as proporções é $m_{ij} = 142$. A: em casa; B: em casa de parentes; C: em casa de amigos; D: na escola; E: no trabalho; F: em *lan-house*; G: biblioteca; H: centros de inclusão digital; I: edição de texto; J: edição de imagem; K: planilha; L: banco de dados; M: acesso a *internet*; N: redes sociais; O: *e-mail*; P: jogos; Q: desenvolvimentos de páginas; R: programação e; S: outras atividades.

APÊNDICE B – Código do Programa R

```

#-----#
# TCC - ANALISE DE VARIÁVEIS CATEGÓRICAS DE RESPOSTA MÚLTIPLA #
#-----#
#           Universidade de Brasília - UnB           #
#           Instituto de Exatas - IE                 #
#           Departamento de Estatística - EST        #
#           2019                                     #
#           #                                         #
#           LEONARDO M. L. PINTO - 16/0033802       #
#           #                                         #
#           ORIENTADOR PROF. DR. GEORGE FREITAS VON BORRIES #
#-----#

#-----#
# Pacotes #
#-----#
# install.packages("MRCV"): Pacote principal para MRCVs
require(MRCV)
# install.packages("vcd"): Gráfico de mosaico
require(vcd)

#-----#
# MENINAS.COMP - Computação também é coisa de menina! (2011-2014) #
#-----#

#-----#
# SUMARIO (Respostas de 2011-2014) #
#-----#
# 1 - Importando o banco de dados
# 2 - Manipulando o banco de dados
#   2.1 - Area de Interesse vs. Local de Uso de Comp.
#   2.2 - Area de Interesse vs. Atividades no Comp.
# 3 - Análise do banco de dados
#   3.1 - Area de Interesse vs. Local de Uso de Comp.
#       3.1.1 - Tabela de contingência marginal

```

```
#      3.1.2 - Tabela de resposta ao item
#      3.1.3 - Gráfico de mosaico
#      3.1.4 - Testes de MMI
#          A - Bootstrap não-paramétrico
#          B - Segunda Ordem Rao-Scott
#          C - Ajuste de Bonferroni
# 3.2 - Area de Interesse vs. Atividades Realizadas no Comp.
#      3.1.1 - Tabela de contingência marginal
#      3.1.2 - Tabela de resposta ao item
#      3.1.3 - Gráfico de mosaico
#      3.1.4 - Testes de MMI
#          A - Bootstrap não-paramétrico
#          B - Segunda Ordem Rao-Scott
#          C - Ajuste de Bonferroni
# 3.3 - Local de Uso do Comp. vs. Atividades Realizadas no Comp.
#      3.3.1 - Exatas
#          3.3.1.1 - Tabela de contingência marginal
#          3.3.1.2 - Tabela de resposta ao item
#          3.3.1.3 - Testes de SPMI
#              A - Bootstrap não-paramétrico
#              B - Segunda Ordem Rao-Scott
#              C - Ajuste de Bonferroni
#      3.3.2 - Exatas
#          3.3.2.1 - Tabela de contingência marginal
#          3.3.2.2 - Tabela de resposta ao item
#          3.3.2.3 - Testes de SPMI
#              A - Bootstrap não-paramétrico
#              B - Segunda Ordem Rao-Scott
#              C - Ajuste de Bonferroni
#      3.3.3 - Exatas
#          3.3.3.1 - Tabela de contingência marginal
#          3.3.3.2 - Tabela de resposta ao item
#          3.3.3.3 - Testes de SPMI
#              A - Bootstrap não-paramétrico
#              B - Segunda Ordem Rao-Scott
#              C - Ajuste de Bonferroni
#-----#
# 1 - Importando o banco de dados #
#-----#
```



```
# Importando o arquivo 'meninas_na_computacao_2019.csv'
ans.path <- file.choose()
ans <- read.csv(ans.path, sep=",", header=T,
               colClasses = rep("factor", 36))

# Primeiras observações
head(ans)

# Resumo dos dados
summary(ans)

#-----#
# 2 - Manipulando o banco de dados #
#-----#
# Precisamos tirar observações com <Year> == 2019
# Precisamos tirar observações com <Gender> = M ou ''
# Precisamos tirar observações com <FOI> = ''

# Retirando as observações
ans2011a2014 <- ans[-which(ans$Year == "2019" | ans$Gender != "F" |
                        ans$FOI == ""),]
summary(ans2011a2014)

# Removendo o fator com 0 observações em <FOI>
FOI2011a2014 <- factor(ans2011a2014$FOI)
summary(FOI2011a2014)

#-----#
# 2.1 - Area de Interesse vs. Local de Uso de Comp. #
#-----#
# Extraindo as colunas relacionadas a pergunta de local de uso de Comp.
localofuse2011a2014 <- ans2011a2014[,18:25]

for(j in 1:ncol(localofuse2011a2014)){
  # Transformando em caracter
  localofuse2011a2014[,j] <- as.character(localofuse2011a2014[,j])

  # Transforma 1's em 0's
  localofuse2011a2014[,j][localofuse2011a2014[,j] == "1"] <- "0"
```

```
# Transforma 3's em 1's
localofuse2011a2014[,j][localofuse2011a2014[,j] == "3"] <- "1"

# Transformando em fator
localofuse2011a2014[,j] <- as.factor(localofuse2011a2014[,j])
}
# Verificando se tudo ocorreu corretamente
summary(localofuse2011a2014)

# Juntando os dados de Area de Interesse e Local de Uso de Comp.
ans.local2011a2014 <- cbind(FOI2011a2014, localofuse2011a2014)

#-----#
# 2.2 - Area de Interesse vs. Atividades no Comp. #
#-----#
# Extraindo as colunas relacionadas a pergunta de atividades no Comp.
useofcomputer2011a2014 <- ans2011a2014[,26:36]

for(j in 1:ncol(useofcomputer2011a2014)){
  # Transformando em caracter
  useofcomputer2011a2014[,j] <- as.character(useofcomputer2011a2014[,j])

  # Transforma 1's em 0's
  useofcomputer2011a2014[,j][useofcomputer2011a2014[,j] == "1"] <- "0"

  # Transforma 3's em 1's
  useofcomputer2011a2014[,j][useofcomputer2011a2014[,j] == "3"] <- "1"

  # Transformando em fator
  useofcomputer2011a2014[,j] <- as.factor(useofcomputer2011a2014[,j])
}
# Verificando se tudo ocorreu corretamente
summary(useofcomputer2011a2014)

# Juntando os dados de Area de Interesse e Atividades no Comp.
ans.uso2011a2014 <- cbind(FOI2011a2014, useofcomputer2011a2014)
```

```
#-----#
# 3 - Análise do banco de dados #
#-----#

#-----#
# 3.1 - Area de Interesse vs. Local de Uso de Comp. #
#-----#

#-----#
# 3.1.1 - Tabela de contingência marginal #
#-----#

tab.marg.local2011a2014 <- marginal.table(ans.local2011a2014, I=1, J=8)

#-----#
# 3.1.2 - Tabela de resposta ao item #
#-----#

item.response.table(ans.local2011a2014, I=1, J=8)

#-----#
# 3.1.3 - Gráfico de mosaico #
#-----#

mosaico.local2011a2014 <- matrix(data = NA, nrow = 3, ncol = 8)

for(j in 1:ncol(tab.marg.local2011a2014[,-1])){
  mosaico.local2011a2014[,j] <- unlist(marginal.table(ans.uso2011a2014,
                                                    I=1, J=8)[,2*j])
}

mosaico.local2011a2014 <- as.data.frame(mosaico.local2011a2014)

# Nomes para o gráfico de mosaico
nome.local <- c("A","B","C","D","E","F","G","H")
nome.areadeinteresse <- c("Exat.", "Bio.S.", "Hum.")
colnames(mosaico.local2011a2014) <- nome.local
rownames(mosaico.local2011a2014) <- nome.areadeinteresse
```

```
# Consertando os dados para o gráfico de mosaico
mosaico.local2011a2014 <- t(mosaico.local2011a2014)
mosaico.local2011a2014 <- t(mosaico.local2011a2014)

# Gráfico de mosaico
mosaic(mosaico.local2011a2014)

#-----#
# 3.1.4 - Testes de MMI #
#-----#

#-----#
# A - Bootstrap não-paramétrico (um pouco demorado!) #
#-----#
r2011a2014.boot.local <- MI.test(ans.local2011a2014, I=1, J=8,
                                type = "boot", print.status = T,
                                add.constant = 0.5)
r2011a2014.boot.local

#-----#
# B - Segunda Ordem Rao-Scott #
#-----#
r2011a2014.rs2.local <- MI.test(ans.local2011a2014, I=1, J=8,
                                type = "rs2", print.status = T,
                                add.constant = 0.5)
r2011a2014.rs2.local

#-----#
# C - Ajuste de Bonferroni #
#-----#
r2011a2014.bon.local <- MI.test(ans.local2011a2014, I=1, J=8,
                                type = "bon", print.status = T,
                                add.constant = 0.5)
r2011a2014.bon.local

#-----#
# 3.2 - Area de Interesse vs. Atividades no Comp. #
#-----#
```

```
#-----#
# 3.2.1 - Tabela de contingência marginal #
#-----#

tab.marg.uso2011a2014 <- marginal.table(ans.uso2011a2014, I=1, J=11)

#-----#
# 3.2.2 - Tabela de resposta ao item #
#-----#

item.response.table(ans.uso2011a2014, I=1, J=11)

#-----#
# 3.2.3 - Gráfico de mosaico #
#-----#

mosaico.uso2011a2014 <- matrix(data = NA, nrow = 3, ncol = 11)

for(j in 1:ncol(tab.marg.uso2011a2014[,-1])){
  mosaico.uso2011a2014[,j] <- unlist(marginal.table(ans.uso2011a2014,
                                                    I=1, J=11)[,2*j])
}

mosaico.uso2011a2014 <- as.data.frame(mosaico.uso2011a2014)

# Nomes para o gráfico de mosaico
nome.uso <- c("I","J","K","L","M","N","O","P","Q","R","S")
nome.areadeinteresse <- c("Exat.", "Bio.S.", "Hum.")
colnames(mosaico.uso2011a2014) <- nome.uso
rownames(mosaico.uso2011a2014) <- nome.areadeinteresse

# Consertando os dados para o gráfico de mosaico
mosaico.uso2011a2014 <-t(mosaico.uso2011a2014)
mosaico.uso2011a2014 <- t(mosaico.uso2011a2014)

# Gráfico de mosaico
mosaic(mosaico.uso2011a2014)
```

```

#-----#
# 3.2.4 - Testes de MMI #
#-----#

#-----#
# A - Bootstrap não-paramétrico (um pouco demorado!) #
#-----#
r2011a2014.boot.uso <- MI.test(ans.uso2011a2014, I=1, J=11, type = "boot",
                             print.status = T, add.constant = 0.5)
r2011a2014.boot.uso

#-----#
# B - Segunda Ordem Rao-Scott #
#-----#
r2011a2014.rs2.uso <- MI.test(ans.uso2011a2014, I=1, J=11, type = "rs2",
                              print.status = T, add.constant = 0.5)
r2011a2014.rs2.uso

#-----#
# C - Ajuste de Bonferroni #
#-----#
r2011a2014.bon.uso <- MI.test(ans.uso2011a2014, I=1, J=11, type = "bon",
                              print.status = T, add.constant = 0.5)
r2011a2014.bon.uso

#-----#
# 3.3 - LOCAL DE USO DO COMP. vs. ATIVIDADES REALIZADAS NO COMP. #
#-----#
# Juntando os dados de local e atividade
locuso2011a2014 <- cbind(FOI2011a2014, localofuse2011a2014,
                       useofcomputer2011a2014)

#-----#
# 3.3.1 - RESPOSTAS GERAIS #
#-----#

Geral2011a2014 <- locuso2011a2014[,-1]

```

```
#-----#
# 3.3.1.1 - Tabela de contingência marginal #
#-----#

marginal.table(Geral2011a2014, I=8, J=11)

#-----#
# 3.3.1.2 - Tabela de resposta ao item #
#-----#

item.response.table(Geral2011a2014, I=8, J=11)

#-----#
# 3.3.1.3 - Testes de SPMI #
#-----#

#-----#
# A - Bootstrap não-paramétrico (um pouco demorado!) #
#-----#
r2011a2014.boot.local.uso.Ge <- MI.test(Geral2011a2014, I=8, J=11,
                                     type = "boot", print.status = T,
                                     add.constant = 0.5)

r2011a2014.boot.local.uso.Ge

#-----#
# B - Segunda Ordem Rao-Scott #
#-----#
r2011a2014.rs2.local.uso.Ge <- MI.test(Exatas2011a2014, I=8, J=11,
                                     type = "rs2", print.status = T,
                                     add.constant = 0.5)

r2011a2014.rs2.local.uso.Ge

#-----#
# C - Ajuste de Bonferroni #
#-----#
r2011a2014.bon.local.uso.Ge <- MI.test(Geral2011a2014, I=8, J=11,
                                     type = "bon", print.status = T,
                                     add.constant = 0.5)

r2011a2014.bon.local.uso.Ge
```

```

#-----#
# 3.3.2 - Area de Interesse: EXATAS #
#-----#

Exatas <- which(locuso2011a2014$FOI2011a2014 == 1)
Exatas2011a2014 <- locuso2011a2014[Exatas,-1]

#-----#
# 3.3.2.1 - Tabela de contingência marginal #
#-----#

marginal.table(Exatas2011a2014, I=8, J=11)

#-----#
# 3.3.2.2 - Tabela de resposta ao item #
#-----#

item.response.table(Exatas2011a2014, I=8, J=11)

#-----#
# 3.3.2.3 - Testes de SPMI #
#-----#

#-----#
# A - Bootstrap não-paramétrico (um pouco demorado!) #
#-----#
r2011a2014.boot.local.uso.Ex <- MI.test(Exatas2011a2014, I=8, J=11,
                                     type = "boot", print.status = T,
                                     add.constant = 0.5)

r2011a2014.boot.local.uso.Ex

#-----#
# B - Segunda Ordem Rao-Scott #
#-----#
r2011a2014.rs2.local.uso.Ex <- MI.test(Exatas2011a2014, I=8, J=11,
                                     type = "rs2", print.status = T,
                                     add.constant = 0.5)

r2011a2014.rs2.local.uso.Ex

```



```
#-----#
# C - Ajuste de Bonferroni #
#-----#
r2011a2014.bon.local.uso.Ex <- MI.test(Exatas2011a2014, I=8, J=11,
                                     type = "bon", print.status = T,
                                     add.constant = 0.5)

r2011a2014.bon.local.uso.Ex

#-----#
# 3.3.3 - Area de Interesse: BIOLÓGICAS #
#-----#

Biologicas <- which(locuso2011a2014$FOI2011a2014 == 2)
Biologicas2011a2014 <- locuso2011a2014[Biologicas,-1]

#-----#
# 3.3.3.1 - Tabela de contingência marginal #
#-----#

marginal.table(Biologicas2011a2014, I=8, J=11)

#-----#
# 3.3.3.2 - Tabela de resposta ao item #
#-----#

item.response.table(Biologicas2011a2014, I=8, J=11)

#-----#
# 3.3.3.3 - Testes de SPMI #
#-----#

#-----#
# A - Bootstrap não-paramétrico (um pouco demorado!) #
#-----#
r2011a2014.boot.local.uso.Bio <- MI.test(Biologicas2011a2014, I=8, J=11,
                                         type = "boot", print.status = T,
                                         add.constant = 0.5)

r2011a2014.boot.local.uso.Bio
```

```

#-----#
# B - Segunda Ordem Rao-Scott #
#-----#
r2011a2014.rs2.local.uso.Bio <- MI.test(Biologicas2011a2014, I=8, J=11,
                                     type = "rs2", print.status = T,
                                     add.constant = 0.5)

r2011a2014.rs2.local.uso.Bio

#-----#
# C - Ajuste de Bonferroni #
#-----#
r2011a2014.bon.local.uso.Bio <- MI.test(Biologicas2011a2014, I=8, J=11,
                                     type = "bon", print.status = T,
                                     add.constant = 0.5)

r2011a2014.bon.local.uso.Bio

#-----#
# 3.3.4 - Area de Interesse: HUMANAS #
#-----#

Humanas <- which(locuso2011a2014$FOI2011a2014 == 3)
Humanas2011a2014 <- locuso2011a2014[Humanas,-1]

#-----#
# 3.3.4.1 - Tabela de contingência marginal #
#-----#

marginal.table(Humanas2011a2014, I=8, J=11)

#-----#
# 3.3.4.2 - Tabela de resposta ao item #
#-----#

item.response.table(Humanas2011a2014, I=8, J=11)

#-----#
# 3.3.4.3 - Testes de SPMI #
#-----#

```

```
#-----#
# A - Bootstrap não-paramétrico (um pouco demorado!) #
#-----#
r2011a2014.boot.local.uso.Hum <- MI.test(Humanas2011a2014, I=8, J=11,
                                         type = "boot", print.status = T,
                                         add.constant = 0.5)

r2011a2014.boot.local.uso.Hum

#-----#
# B - Segunda Ordem Rao-Scott #
#-----#
r2011a2014.rs2.local.uso.Hum <- MI.test(Humanas2011a2014, I=8, J=11,
                                         type = "rs2", print.status = T,
                                         add.constant = 0.5)

r2011a2014.rs2.local.uso.Hum

#-----#
# C - Ajuste de Bonferroni #
#-----#
r2011a2014.bon.local.uso.Hum <- MI.test(Humanas2011a2014, I=8, J=11,
                                         type = "bon", print.status = T,
                                         add.constant = 0.5)

r2011a2014.bon.local.uso.Hum

#-----#
# SUMARIO (Respostas de 2019) #
#-----#
# 4 - Manipulando o banco de dados
#   4.1 - Area de Interesse vs. Local de Uso de Comp.
#   4.2 - Area de Interesse vs. Atividades no Comp.
# 5 - Análise do banco de dados
#   5.1 - Area de Interesse vs. Local de Uso de Comp.
#     5.1.1 - Tabela de contingência marginal
#     5.1.2 - Tabela de resposta ao item
#     5.1.3 - Gráfico de mosaico
#     5.1.4 - Testes de MMI
#       A - Bootstrap não-paramétrico
#       B - Segunda Ordem Rao-Scott
```

```
#           C - Ajuste de Bonferroni
# 5.2 - Area de Interesse vs. Atividades Realizadas no Comp.
#   5.2.1 - Tabela de contingência marginal
#   5.2.2 - Tabela de resposta ao item
#   5.2.3 - Gráfico de mosaico
#   5.2.4 - Testes de MMI
#     A - Bootstrap não-paramétrico
#     B - Segunda Ordem Rao-Scott
#     C - Ajuste de Bonferroni
# 5.3 - Local de Uso do Comp. vs. Atividades Realizadas no Comp.
#   5.3.1 - Geral
#     5.3.1.1 - Tabela de contingência marginal
#     5.3.1.2 - Tabela de resposta ao item
#     5.3.1.3 - Testes de SPMI
#       A - Bootstrap não-paramétrico
#       B - Segunda Ordem Rao-Scott
#       C - Ajuste de Bonferroni
#   5.3.2 - Exatas
#     5.3.2.1 - Tabela de contingência marginal
#     5.3.2.2 - Tabela de resposta ao item
#     5.3.2.3 - Testes de SPMI
#       A - Bootstrap não-paramétrico
#       B - Segunda Ordem Rao-Scott
#       C - Ajuste de Bonferroni
#   5.3.3 - Exatas
#     5.3.3.1 - Tabela de contingência marginal
#     5.3.3.2 - Tabela de resposta ao item
#     5.3.3.3 - Testes de SPMI
#       A - Bootstrap não-paramétrico
#       B - Segunda Ordem Rao-Scott
#       C - Ajuste de Bonferroni
#   5.3.4 - Exatas
#     5.3.4.1 - Tabela de contingência marginal
#     5.3.4.2 - Tabela de resposta ao item
#     5.3.4.3 - Testes de SPMI
#       A - Bootstrap não-paramétrico
#       B - Segunda Ordem Rao-Scott
#       C - Ajuste de Bonferroni
# 6 - Preparacao de dados para modelo log-linear
```

```
# 6.1 - Juntando niveis para LOCAL DE USO DO Comp.
# 6.1.1 - Em casa (A)
# 6.1.2 - Em casa de amigos (B) e em casa de parentes (C)
# 6.1.3 - Na escola (D) e no trabalho (E)
# 6.1.4 - Lan-house (F), bibliot. (G) e centr. de incl. digit. (H)
# 6.2 - Juntando niveis para ATIVIDADES REALIZADAS NO Comp.
# 6.2.1 - Edição de textos (I) e imagens (J)
# 6.2.2 - Planilhas (K) e banco de dados (L)
# 6.2.3 - Internet (M), redes sociais (N) e e-mail (O)
# 6.2.4 - Jogos (P), desenv. de pag. (Q), program. (R) e outr. (S)
# 6.3 - Juntando as respostas
# 7 - Modelos
# 7.1 - Ajuste
# 7.2 - Comparando modelos
# 7.2.1 - X vs. outros
# 7.2.2 - Y vs. outros
# 7.2.3 - XY vs. outros
# 7.2.4 - SPMI vs. outros
# 7.2.5 - Homogeneo vs. outros

#-----#
# 4 - Manipulando o banco de dados #
#-----#

# Precisamos tirar observações com <Year> != 2019
# Precisamos tirar observações com <Gender> = M ou ''
# Precisamos tirar observações com <FOI> = ''

# Retirando as observações
ans2019 <- ans[-which(ans$Year != "2019" | ans$Gender != "F" |
                    ans$FOI == ""),]
summary(ans2019)

# Removendo o fator com 0 observações em <FOI>
FOI2019 <- factor(ans2019$FOI)
summary(FOI2019)
```

```

#-----#
# 4.1 - Area de Interesse vs. Local de Uso de Comp. #
#-----#
# Extraindo as colunas relacionadas a pergunta de local de uso de Comp.
localofuse2019 <- ans2019[,18:25]

for(j in 1:ncol(localofuse2019)){
  # Transformando em caracter
  localofuse2019[,j] <- as.character(localofuse2019[,j])

  # Transforma 1's em 0's
  localofuse2019[,j][localofuse2019[,j] == "1"] <- "0"

  # Transforma 3's em 1's
  localofuse2019[,j][localofuse2019[,j] == "3"] <- "1"

  # Transformando em fator
  localofuse2019[,j] <- as.factor(localofuse2019[,j])
}
# Verificando se tudo ocorreu corretamente
summary(localofuse2019)

# Juntando os dados de Area de Interesse e Local de Uso de Comp.
ans.local2019 <- cbind(FOI2019, localofuse2019)

#-----#
# 4.2 - Area de Interesse vs. Atividades no Comp. #
#-----#

# Extraindo as colunas relacionadas a pergunta de atividades no Comp.
useofcomputer2019 <- ans2019[,26:36]

for(j in 1:ncol(useofcomputer2019)){
  # Transformando em caracter
  useofcomputer2019[,j] <- as.character(useofcomputer2019[,j])

  # Transforma 1's em 0's
  useofcomputer2019[,j][useofcomputer2019[,j] == "1"] <- "0"
}

```

```
# Transforma 3's em 1's
useofcomputer2019[,j][useofcomputer2019[,j] == "3"] <- "1"

# Transformando em fator
useofcomputer2019[,j] <- as.factor(useofcomputer2019[,j])
}
# Verificando se tudo ocorreu corretamente
summary(useofcomputer2019)

# Juntando os dados de Area de Interesse e Atividades no Comp.
ans.uso2019 <- cbind(FOI2019, useofcomputer2019)

#-----#
# 5 - Análise do banco de dados #
#-----#

#-----#
# 5.1 - Area de Interesse vs. Local de Uso de Comp. #
#-----#

#-----#
# 5.1.1 - Tabela de contingência marginal #
#-----#

tab.marg.local2019 <- marginal.table(ans.local2019, I=1, J=8)

#-----#
# 5.1.2 - Tabela de resposta ao item #
#-----#

item.response.table(ans.local2019, I=1, J=8)

#-----#
# 5.1.3 - Gráfico de mosaico #
#-----#

mosaico.local2019 <- matrix(data = NA, nrow = 3, ncol = 8)
```

```

for(j in 1:ncol(tab.marg.local2019[,-1])){
  mosaico.local2019[,j] <- unlist(marginal.table(ans.uso2019,
                                             I=1, J=8)[,2*j])
}

mosaico.local2019 <- as.data.frame(mosaico.local2019)

# Nomes para o gráfico de mosaico
nome.local <- c("A","B","C","D","E","F","G","H")
nome.areadeinteresse <- c("Exat.", "Bio.S.", "Hum.")
colnames(mosaico.local2019) <- nome.local
rownames(mosaico.local2019) <- nome.areadeinteresse

# Consertando os dados para o gráfico de mosaico
mosaico.local2019 <- t(mosaico.local2019)
mosaico.local2019 <- t(mosaico.local2019)

# Gráfico de mosaico
mosaic(mosaico.local2019)

#-----#
# 5.1.4 - Testes de MMI #
#-----#

#-----#
# A - Bootstrap não-paramétrico (um pouco demorado!) #
#-----#
r2019.boot.local <- MI.test(ans.local2019, I=1, J=8, type = "boot",
                           print.status = T, add.constant = 0.5)
r2019.boot.local

#-----#
# B - Segunda Ordem Rao-Scott #
#-----#
r2019.rs2.local <- MI.test(ans.local2019, I=1, J=8, type = "rs2",
                           print.status = T, add.constant = 0.5)
r2019.rs2.local

```



```
#-----#
# C - Ajuste de Bonferroni #
#-----#
r2019.bon.local <- MI.test(ans.local2019, I=1, J=8, type = "bon",
                           print.status = T, add.constant = 0.5)
r2019.bon.local

#-----#
# 5.2 - Area de Interesse vs. Atividades no Comp. #
#-----#

#-----#
# 5.2.1 - Tabela de contingência marginal #
#-----#

tab.marg.uso2019 <- marginal.table(ans.uso2019, I=1, J=11)

#-----#
# 5.2.2 - Tabela de resposta ao item #
#-----#

item.response.table(ans.uso2019, I=1, J=11)

#-----#
# 5.2.3 - Gráfico de mosaico #
#-----#

mosaico.uso2019 <- matrix(data = NA, nrow = 3, ncol = 11)

for(j in 1:ncol(tab.marg.uso2019[,-1])){
  mosaico.uso2019[,j] <- unlist(marginal.table(ans.uso2019,
                                              I=1, J=11)[,2*j])
}

mosaico.uso2019 <- as.data.frame(mosaico.uso2019)

# Nomes para o gráfico de mosaico
nome.uso <- c("I","J","K","L","M","N","O","P","Q","R","S")
nome.areadeinteresse <- c("Exat.", "Bio.S.", "Hum.")
```

```
colnames(mosaico.uso2019) <- nome.uso
rownames(mosaico.uso2019) <- nome.areadeinteresse

# Consertando os dados para o gráfico de mosaico
mosaico.uso2019 <-t(mosaico.uso2019)
mosaico.uso2019 <- t(mosaico.uso2019)

# Gráfico de mosaico
mosaic(mosaico.uso2019)

#-----#
# 5.2.4 - Testes de MMI #
#-----#

#-----#
# A - Bootstrap não-paramétrico (um pouco demorado!) #
#-----#
r2019.boot.uso <- MI.test(ans.uso2019, I=1, J=11, type = "boot",
                        print.status = T, add.constant = 0.5)
r2019.boot.uso

#-----#
# B - Segunda Ordem Rao-Scott #
#-----#
r2019.rs2.uso <- MI.test(ans.uso2019, I=1, J=11, type = "rs2",
                        print.status = T, add.constant = 0.5)
r2019.rs2.uso

#-----#
# C - Ajuste de Bonferroni #
#-----#
r2019.bon.uso <- MI.test(ans.uso2019, I=1, J=11, type = "bon",
                        print.status = T, add.constant = 0.5)
r2019.bon.uso

#-----#
# 5.3 - LOCAL DE USO DO COMP. vs. ATIVIDADES REALIZADAS NO COMP. #
#-----#
```

```
# Juntando os dados de local e atividade
locuso2019 <- cbind(FOI2019, localofuse2019, useofcomputer2019)

#-----#
# 5.3.1 - RESPOSTAS GERAIS #
#-----#

Geral2019 <- locuso2019[,-1]

#-----#
# 5.3.1.1 - Tabela de contingência marginal #
#-----#

marginal.table(Geral2019, I=8, J=11)

#-----#
# 5.3.2.2 - Tabela de resposta ao item #
#-----#

item.response.table(Geral2019, I=8, J=11)

#-----#
# 5.3.2.3 - Testes de SPMI #
#-----#

#-----#
# A - Bootstrap não-paramétrico (um pouco demorado!) #
#-----#
r2019.boot.local.uso.Ge <- MI.test(Geral2019, I=8, J=11,
                                type = "boot", print.status = T,
                                add.constant = 0.5)
r2019.boot.local.uso.Ge

#-----#
# B - Segunda Ordem Rao-Scott #
#-----#
r2019.rs2.local.uso.Ge <- MI.test(Exatas2019, I=8, J=11,
                                type = "rs2", print.status = T,
                                add.constant = 0.5)
```

```

r2019.rs2.local.uso.Ge

#-----#
# C - Ajuste de Bonferroni #
#-----#
r2019.bon.local.uso.Ge <- MI.test(Geral2019, I=8, J=11,
                                type = "bon", print.status = T,
                                add.constant = 0.5)

r2019.bon.local.uso.Ge

#-----#
# 5.3.2 - Area de Interesse: EXATAS #
#-----#

Exatas2019 <- locuso2019[which(locuso2019$FOI2019 == 1),-1]

#-----#
# 5.3.2.1 - Tabela de contingência marginal #
#-----#

marginal.table(Exatas2019, I=8, J=11)

#-----#
# 5.3.2.2 - Tabela de resposta ao item #
#-----#

item.response.table(Exatas2019, I=8, J=11)

#-----#
# 5.3.2.3 - Testes de SPMI #
#-----#

#-----#
# A - Bootstrap não-paramétrico (um pouco demorado!) #
#-----#
r2019.boot.local.uso.Ex <- MI.test(Exatas2019, I=8, J=11,
                                type = "boot", print.status = T,
                                add.constant = 0.5)

r2019.boot.local.uso.Ex

```

```
#-----#
# B - Segunda Ordem Rao-Scott #
#-----#
r2019.rs2.local.uso.Ex <- MI.test(Exatas2019, I=8, J=11,
                                type = "rs2", print.status = T,
                                add.constant = 0.5)
r2019.rs2.local.uso.Ex

#-----#
# C - Ajuste de Bonferroni #
#-----#
r2019.bon.local.uso.Ex <- MI.test(Exatas2019, I=8, J=11,
                                type = "bon", print.status = T,
                                add.constant = 0.5)
r2019.bon.local.uso.Ex

#-----#
# 5.3.3 - Area de Interesse: BIOLÓGICAS #
#-----#

Biologicas2019 <- locuso2019[which(locuso2019$FOI2019 == 2),-1]

#-----#
# 5.3.3.1 - Tabela de contingência marginal #
#-----#

marginal.table(Biologicas2019, I=8, J=11)

#-----#
# 5.3.3.2 - Tabela de resposta ao item #
#-----#

item.response.table(Biologicas2019, I=8, J=11)

#-----#
# 5.3.3.3 - Testes de SPMI #
#-----#
```

```

#-----#
# A - Bootstrap não-paramétrico (um pouco demorado!) #
#-----#
r2019.boot.local.uso.Bio <- MI.test(Biologicas2019, I=8, J=11,
                                   type = "boot", print.status = T,
                                   add.constant = 0.5)

r2019.boot.local.uso.Bio

#-----#
# B - Segunda Ordem Rao-Scott #
#-----#
r2019.rs2.local.uso.Bio <- MI.test(Biologicas2019, I=8, J=11,
                                   type = "rs2", print.status = T,
                                   add.constant = 0.5)

r2019.rs2.local.uso.Bio

#-----#
# C - Ajuste de Bonferroni #
#-----#
r2019.bon.local.uso.Bio <- MI.test(Biologicas2019, I=8, J=11,
                                   type = "bon", print.status = T,
                                   add.constant = 0.5)

r2019.bon.local.uso.Bio

#-----#
# 5.3.4 - Area de Interesse: HUMANAS #
#-----#

Humanas2019 <- locuso2019[which(locuso2019$FOI2019 == 3),-1]

#-----#
# 5.3.4.1 - Tabela de contingência marginal #
#-----#

marginal.table(Humanas2019, I=8, J=11)

#-----#
# 5.3.4.2 - Tabela de resposta ao item #
#-----#

```

```
item.response.table(Humanas2019, I=8, J=11)

#-----#
# 5.3.4.3 - Testes de SPMI #
#-----#

#-----#
# A - Bootstrap não-paramétrico (um pouco demorado!) #
#-----#
r2019.boot.local.uso.Hum <- MI.test(Humanas2019, I=8, J=11,
                                   type = "boot", print.status = T,
                                   add.constant = 0.5)

r2019.boot.local.uso.Hum

#-----#
# B - Segunda Ordem Rao-Scott #
#-----#
r2019.rs2.local.uso.Hum <- MI.test(Humanas2019, I=8, J=11,
                                   type = "rs2", print.status = T,
                                   add.constant = 0.5)

r2019.rs2.local.uso.Hum

#-----#
# C - Ajuste de Bonferroni #
#-----#
r2019.bon.local.uso.Hum <- MI.test(Humanas2019, I=8, J=11,
                                   type = "bon", print.status = T,
                                   add.constant = 0.5)

r2019.bon.local.uso.Hum

#-----#
# 6 - Preparacao de dados para modelo log-linear #
#-----#

#-----#
# 6.1 - Juntando niveis para LOCAL DE USO DO COMP. #
#-----#
```

```

#-----#
# 6.1.1 - Em casa (A) #
#-----#

ansMRCV1.1.2019 <- ans.local2019$UCH

#-----#
# 6.1.2 - Em casa de amigos (B) e em casa de parentes (C) #
#-----#

B <- as.numeric(as.character(ans.local2019$UCRH))
C <- as.numeric(as.character(ans.local2019$UCFH))
ansMRCV.casa.2019 <- cbind(data.frame(B), data.frame(C))

# Se a soma da linha e igual a 0, nenhuma respondente usa o PC em casa de
# amigos ou em casa de parentes, logo atribua 0. C.C., atribua 1.
ansMRCV1.2.2019 <- ifelse(rowSums(ansMRCV.casa.2019) == 0,
                          ansMRCV1.2.2019 <- 0,
                          ansMRCV1.2.2019 <- 1)

# Transformando em fator
ansMRCV1.2.2019 <- as.factor(ansMRCV1.2.2019)

# Verificando
ansMRCV1.2.2019

#-----#
# 6.1.3 - Na escola (D) e no trabalho (E) #
#-----#

D <- as.numeric(as.character(ans.local2019$UCS))
E <- as.numeric(as.character(ans.local2019$UCW))
ansMRCV.trab.2019 <- cbind(data.frame(D), data.frame(E))

# Se a soma da linha e igual a 0, nenhuma respondente usa o PC na escola
# ou no trabalho, logo atribua 0. C.C., atribua 1.
ansMRCV1.3.2019 <- ifelse(rowSums(ansMRCV.trab.2019) == 0,
                          ansMRCV1.3.2019 <- 0,
                          ansMRCV1.3.2019 <- 1)

```



```
# Transformando em fator
ansMRCV1.3.2019 <- as.factor(ansMRCV1.3.2019)

# Verificando
ansMRCV1.3.2019

#-----#
# 6.1.4 - Lan-house (F), biblioteca (G) e centr. de inclusao digit. (H) #
#-----#

FF <- as.numeric(as.character(ans.local2019$UCLH))
G <- as.numeric(as.character(ans.local2019$UCAL))
H <- as.numeric(as.character(ans.local2019$UCDIC))
ansMRCV.outr.2019 <- cbind(data.frame(FF), data.frame(G), data.frame(H))

# Se a soma da linha e igual a 0, nenhuma respondente usa o PC em
# lan-house, biblioteca ou em CdID, logo atribua 0. C.C., atribua 1.
ansMRCV1.4.2019 <- ifelse(rowSums(ansMRCV.outr.2019) == 0,
                          ansMRCV1.4.2019 <- 0,
                          ansMRCV1.4.2019 <- 1)

# Transformando em fator
ansMRCV1.4.2019 <- as.factor(ansMRCV1.4.2019)

# Verificando
ansMRCV1.4.2019

#-----#
# 6.2 - Juntando niveis para ATIVIDADES REALIZADAS NO COMP. #
#-----#

#-----#
# 6.2.1 - Edição de textos (I) e imagens (J) #
#-----#

I <- as.numeric(as.character(ans.uso2019$HUTE))
J <- as.numeric(as.character(ans.uso2019$HUIE))
```

```
ansMRCV.edicao.2019 <- cbind(data.frame(I), data.frame(J))

# Se a soma da linha e igual a 0, nenhuma respondente usa o PC para
# editar textos ou imagens, logo atribua 0. C.C., atribua 1.
ansMRCV2.1.2019 <- ifelse(rowSums(ansMRCV.edicao.2019) == 0,
                          ansMRCV2.1.2019 <- 0,
                          ansMRCV2.1.2019 <- 1)

# Transformando em fator
ansMRCV2.1.2019 <- as.factor(ansMRCV2.1.2019)

# Verificando
ansMRCV2.1.2019

#-----#
# 6.2.2 - Planilhas (K) e banco de dados (L) #
#-----#

K <- as.numeric(as.character(ans.uso2019$HUSP))
L <- as.numeric(as.character(ans.uso2019$HUDB))

ansMRCV.dados.2019 <- cbind(data.frame(K), data.frame(L))

# Se a soma da linha e igual a 0, nenhuma respondente usa o PC para
# planilhas ou banco de dados, logo atribua 0. C.C., atribua 1.
ansMRCV2.2.2019 <- ifelse(rowSums(ansMRCV.dados.2019) == 0,
                          ansMRCV2.2.2019 <- 0,
                          ansMRCV2.2.2019 <- 1)

# Transformando em fator
ansMRCV2.2.2019 <- as.factor(ansMRCV2.2.2019)

# Verificando
ansMRCV2.2.2019

#-----#
# 6.2.3 - Internet (M), redes sociais (N) e e-mail (O) #
#-----#
```

```
M <- as.numeric(as.character(ans.uso2019$HUIT))
N <- as.numeric(as.character(ans.uso2019$HUSN))
O <- as.numeric(as.character(ans.uso2019$HUEM))

ansMRCV.net.2019 <- cbind(data.frame(M), data.frame(N), data.frame(O))

# Se a soma da linha e igual a 0, nenhuma respondente usa o PC para
# internet, redes sociais ou e-mail, logo atribua 0. C.C., atribua 1.
ansMRCV2.3.2019 <- ifelse(rowSums(ansMRCV.net.2019) == 0,
                          ansMRCV2.3.2019 <- 0,
                          ansMRCV2.3.2019 <- 1)

# Transformando em fator
ansMRCV2.3.2019 <- as.factor(ansMRCV2.3.2019)

# Verificando
ansMRCV2.3.2019

#-----#
# 6.2.4 - Jogos (P), desenv. de pag. (Q), programacao (R) e outr. (S) #
#-----#

P <- as.numeric(as.character(ans.uso2019$HUGM))
Q <- as.numeric(as.character(ans.uso2019$HUCW))
R <- as.numeric(as.character(ans.uso2019$HUFDD))
S <- as.numeric(as.character(ans.uso2019$HUOS))

ansMRCV.outros.2019 <- cbind(data.frame(P), data.frame(Q),
                             data.frame(R), data.frame(S))

# Se a soma da linha e igual a 0, nenhuma respondente usa o PC para jogos,
# desenv. paginas, programar ou outros, logo atribua 0. C.C., atribua 1.
ansMRCV2.4.2019 <- ifelse(rowSums(ansMRCV.outros.2019) == 0,
                          ansMRCV2.4.2019 <- 0,
                          ansMRCV2.4.2019 <- 1)

# Transformando em fator
ansMRCV2.4.2019 <- as.factor(ansMRCV2.4)
```

```

# Verificando
ansMRCV2.4.2019

#-----#
# 6.3 - Juntando as respostas #
#-----#

ansMRCV.novo.2019 <- data.frame(ansMRCV1.1.2019, ansMRCV1.2.2019,
                               ansMRCV1.3.2019, ansMRCV1.4.2019,
                               ansMRCV2.1.2019, ansMRCV2.2.2019,
                               ansMRCV2.3.2019, ansMRCV2.4.2019)

# Verificando
ansMRCV.novo.2019

#-----#
# 7 - MODELOS #
#-----#

#-----#
# 7.1 - Ajuste #
#-----#

#-----#
# "w.main", "y.main", "wy.main", "spmi", "homogeneous", "saturated" #
#-----#

wmain.2019 <- genloglin(data = ansMRCV.novo.2019, I = 4, J = 4,
                      model = "w.main", B = 1999, print.status = T)
ymain.2019 <- genloglin(data = ansMRCV.novo.2019, I = 4, J = 4,
                      model = "y.main", B = 1999, print.status = T)
wymain.2019 <- genloglin(data = ansMRCV.novo.2019, I = 4, J = 4,
                       model = "wy.main", B = 1999, print.status = T)
spmi.2019 <- genloglin(data = ansMRCV.novo.2019, I = 4, J = 4,
                      model = "spmi", B = 1999, print.status = T)
homo.2019 <- genloglin(data = ansMRCV.novo.2019, I = 4, J = 4,
                      model = "homogeneous", B = 1999, print.status = T)
sat.2019 <- genloglin(data = ansMRCV.novo.2019, I = 4, J = 4,
                     model = "saturated", B = 1999, print.status = T)

```

```
#-----#
# 7.2 - Comparacao de modelos #
#-----#

#-----#
# 7.2.1 - X vs. outros #
#-----#
anova(object = wmain.2019, model.HA = "y.main", type = "all")
anova(object = wmain.2019, model.HA = "wy.main", type = "all")
anova(object = wmain.2019, model.HA = count ~ -1 + W:Y + wi %in% W:Y +
      yj %in% W:Y, type = "all")
anova(object = wmain.2019, model.HA = "homogeneous", type = "all")
anova(object = wmain.2019, model.HA = "saturated", type = "all")

#-----#
# 7.2.2 - Y vs. outros #
#-----#
anova(object = ymain.2019, model.HA = "w.main", type = "all")
anova(object = ymain.2019, model.HA = "wy.main", type = "all")
anova(object = ymain.2019, model.HA = count ~ -1 + W:Y + wi %in% W:Y +
      yj %in% W:Y, type = "all")
anova(object = ymain.2019, model.HA = "homogeneous", type = "all")
anova(object = ymain.2019, model.HA = "saturated", type = "all")

#-----#
# 7.2.3 - XY vs. outros #
#-----#
anova(object = wymain.2019, model.HA = "w.main", type = "all")
anova(object = wymain.2019, model.HA = "y.main", type = "all")
anova(object = wymain.2019, model.HA = count ~ -1 + W:Y + wi %in% W:Y +
      yj %in% W:Y, type = "all")
anova(object = wymain.2019, model.HA = "homogeneous", type = "all")
anova(object = wymain.2019, model.HA = "saturated", type = "all")

#-----#
# 7.2.4 - SPMI vs. outros #
#-----#
anova(object = spmi.2019, model.HA = "w.main", type = "all")
```

```
anova(object = spmi.2019, model.HA = "y.main", type = "all")
anova(object = spmi.2019, model.HA = "wy.main", type = "all")
anova(object = spmi.2019, model.HA = "homogeneous", type = "all")
anova(object = spmi.2019, model.HA = "saturated", type = "all")

#-----#
# 7.2.5 - Homogeneo vs. outros #
#-----#
anova(object = homo.2019, model.HA = "w.main", type = "all")
anova(object = homo.2019, model.HA = "y.main", type = "all")
anova(object = homo.2019, model.HA = "wy.main", type = "all")
anova(object = homo.2019, model.HA = count ~ -1 + W:Y + wi %in% W:Y +
      yj %in% W:Y, type = "all")
anova(object = homo.2019, model.HA = "saturated", type = "all")
```