

Larissa Caroline Lino Rosa

Aplicação de visualização em gráficos mosaico com dados psicossociais

Uberlândia

2018

Larissa Caroline Lino Rosa

Aplicação de visualização em gráficos mosaico com dados psicossociais

Trabalho de conclusão de curso apresentado ao Instituto de Psicologia da Universidade Federal de Uberlândia, como requisito parcial à obtenção do Título de Bacharel em Psicologia.

Orientador: João Fernando Rech Wachelke

Uberlândia

2018

Larissa Caroline Lino Rosa

Aplicação de visualização em gráficos mosaico com dados psicossociais

Trabalho de conclusão de curso apresentado ao Instituto de Psicologia da Universidade Federal de Uberlândia, como requisito parcial à obtenção do Título de Bacharel em Psicologia.

Orientador: João Fernando Rech Wachelke

Banca Examinadora

Uberlândia, 30 de novembro de 2018

Prof. Dr. João Fernando Rech Wachelke
Universidade Federal de Uberlândia – Uberlândia MG

Prof^a. Dr^a. Renata Fabiana Pegoraro
Universidade Federal de Uberlândia – Uberlândia MG

Bacharel Rafael Camilo Gonçalves
Universidade Federal de Uberlândia – Uberlândia MG

Uberlândia

2018

Resumo

Este trabalho tem como objetivo apresentar exemplos de aplicação com dados psicossociais de uma técnica de visualização pertinente para a estatística descritiva de variáveis categóricas, chamada gráficos mosaico. Inicialmente, será abordado o conceito de estatística descritiva, que é um dos dois grandes campos, da subdivisão da Estatística, e que visa representar de maneira mais organizada as informações coletadas em uma pesquisa, de acordo com as variáveis selecionadas, através de tabelas e gráficos. Variáveis essas que terão suas discrepâncias e particularidades salientadas, deixando claro que, em casos de pesquisas psicossociais, abordagem trabalhada neste artigo, as mais utilizadas são as categóricas, já que dizem respeito a qualidade ou característica daquilo que pretende-se observar. E através de exemplos, será mostrado como as informações obtidas dessas variáveis podem ser representadas, utilizando as Tabelas de Contingências. Após essa contextualização teórica, o texto mostrará que existe métodos que facilitam a leitura de tabelas de contingências, principalmente, em casos que utiliza-se mais de duas variáveis, deixando a compreensão e visualização dos dados menos complexa, por meio dos gráficos mosaicos.

Palavras-chaves: estatística descritiva, variáveis categóricas, tabela de contingência, gráfico mosaico

Abstract

This work aims to present application examples with psychosocial data of a visualization technique relevant to the descriptive statistics of categorical variables, called Mosaic displays. Initially, it will be discussed the concept of descriptive statistics, which is one of the two major fields, the subdivision of statistics, and which aims to represent a more organized information gathered in a search, according to selected variables, using charts and graphs. These variables will have their discrepancies and particularities highlighted, making it clear that in cases of psychosocial research, crafted approach in this article, the most used are the categorical, since they relate to quality or characteristic what you want to see. Through examples, will be shown as the information obtained of these variables can be represented using Contingency tables. After this theoretical contextualization, the text will show that there are methods that make it easy to read tables of contingencies, especially in cases that use more than two variables, leaving the understanding and visualization of less complex data, through Mosaic displays.

Key-words: descriptive statistics, categorical variables, contingency table, mosaic displays.

Introdução

O presente trabalho apresenta exemplos de aplicação com dados psicossociais de uma técnica de visualização pertinente para a estatística descritiva de variáveis categóricas. Inicialmente, portanto, cabe situar a área e suas diferenças com outros campos da estatística, bem como alguns conceitos essenciais.

Gonçalves (1977) traz que a Estatística é a ciência que estuda dados e examina as informações contidas neles, por meio da coleta, análise e interpretação dos mesmos. Tornou-se uma disciplina científica em meados do século XVII, através do professor de Filosofia, Medicina e Política da Universidade de Helmstadt, Herman Conring, que inaugurou um curso de Ciência Política, em 1660, cujo objetivo era analisar e descrever as questões intrínsecas ao Estado.

Após esse pontapé, houve sucessores que contribuíram para o desenvolvimento da teoria e proporcionaram a sistematização da mesma, que atualmente faz parte do cotidiano e é indispensável nas pesquisas científicas. A estatística se divide em dois grandes campos: a estatística descritiva e a estatística indutiva (Gonçalves, 1977).

A estatística indutiva, também conhecida como inferência estatística, tem como objetivo caracterizar o todo a partir de uma população, coletando e analisando os dados através de comparações, generalizações, ligações lógicas, a fim de estabelecer indicações de vantagens e desvantagens dos resultados obtidos (Cabral & Guimarães, 1997).

Já a estatística descritiva, ou análise exploratória de dados, segundo Cabral e Guimarães (1997) busca representar de maneira organizada e compreensível a informação contida nos dados, através de sua classificação e síntese, por meio de tabelas, gráficos e medidas. Mirshawka (1972) acrescenta que esse modelo de análise faz com que um número, por si só, descreva características de um determinado conjunto de dados.

Os dados são as informações retiradas de alguma variável, que são características

observáveis (por exemplo: idade, sexo, estado civil, escolaridade) de uma população. Podem ser caracterizados como univariados, bivariados e multivariados (Tiboni, 2010).

Segundo Jelihovschi (2014), dados univariados são aqueles cuja análise é pautada em apenas uma variável e no resultado da medida da mesma sobre a amostra utilizada. Dados bivariados são o resultado da medição de duas variáveis diferentes em um mesmo conjunto de objetos. Já os multivariados dizem respeito a quando várias variáveis são medidas em cada objeto da amostra selecionada (Jelihovschi, 2014).

Gil (2008), de maneira mais detalhada, traz como se dá esses três tipos de análises na estatística descritiva: univariada, bivariada e multivariada. Na primeira, os objetivos são: caracterizar o que é típico em um grupo, de modo que o represente como um todo (medidas de tendência central), por meio da mensuração de sua média, mediana e moda; identificar o grau de semelhança ou diferença entre os sujeitos, o que conseqüentemente, leva aos extremos que podem ser encontrados, as exceções, que não se assemelham as características gerais, calculando a amplitude, desvio médio, desvio padrão e o desvio quartílico da amostra; e verificar a distribuição da amostra em relação a variável estabelecida.

Na análise bivariada o objetivo é verificar se há relação entre duas variáveis, a fim de testar hipóteses, por meio de testes de correlação, os quais vão indicar correlações positivas, negativas ou inexistentes entre ambas. Por fim, a análise multivariada é uma extensão das anteriores, sendo caracterizada por elaborar uma matriz de dados, na qual resume e reduza os dados das correlações de várias variáveis; analisar, isoladamente, a relação de uma única variável com as demais; agregar objetos (pessoas, produtos, por exemplo) que possuem características semelhantes em relação a uma variável pré determinada; e classificar elementos dos grupos da população (Gil, 2008).

Como relatado acima, os dados são extraídos de uma determinada amostra, que é um conjunto de elementos, indivíduos, por exemplo, retirados de uma população, que é o todo

(Stevenson, 1981). Jelihovschi (2014) relata que as amostras podem ser selecionadas de três maneiras: causal simples, estratificada e sistemática. Amostra causal simples, é quando qualquer indivíduo tem chance de ser incluído no conjunto que será selecionado, através de um sorteio, por exemplo, sem que haja nenhum pré-requisito. Amostra estratificada é quando a população é dividida em subgrupos com itens semelhantes. E na amostra sistemática, estabelece-se uma sequência na qual os itens serão definidos de maneira aleatória. Exemplo: em uma sala de aula que contem 50 alunos, tendo como referência o número da lista de chamada, de 2 em 2, serão selecionados estudantes que farão parte da pesquisa, ou seja, os que ocupam a posição 2, 4, 6, 8...

Assim como as amostras, outro conceito basilar em uma pesquisa consiste nas variáveis. Dancey e Reidy (2008) trazem que elas são primordiais na pesquisa científica, pois são o que se pretende medir. São fatores que podem assumir valores passíveis de mensuração e registro e que variam dependendo da situação ou pessoa.

Segundo Tiboni (2010), as variáveis podem ser qualitativas (categóricas) ou quantitativas. As variáveis quantitativas dizem respeito a expressões numéricas, tais como, peso, idade, altura, etc. Elas podem ser contínuas ou discretas. A primeira, diz respeito a contagem, a valores inteiros, por exemplo: quantidade de ovos dentro da caixa (6, 12, 30). A segunda, assume valores decimais, é uma medida numérica, como altura, tempo de espera na fila etc (Tiboni 2010).

As qualitativas ou categóricas expressam qualidade ou característica (ex: sexo, escolaridade). Podem ser subdivididas entre nominais e ordinais. As ordinais estabelecem uma ordem, como escolaridade, por exemplo, que possui ensino fundamental, médio e superior. As nominais são aquelas que nomeiam as modalidades da variável sem correlação com alguma ordem, por exemplo, a variável sexo, pode incluir modalidades como masculino ou feminino (Tiboni 2010).

Campos (2000) relata que as variáveis nominais estão fortemente presentes em pesquisas cujas perguntas de um questionário admitem apenas respostas pré estabelecidas, por exemplo: concordo totalmente, discordo totalmente, não concordo/nem discordo. Dessa maneira, os dados que até então são nominais, a partir do momento que se considera a quantidade de resposta de concordância, discordância e as neutras, passam a ter uma expressão numérica. Ou seja, mesmo quando o estudo envolve dados de natureza qualitativa, para que a análise estatística seja realizada, é necessário transformá-los em quantitativos.

Kellner e Nick (1977) em suas concepções sobre estatística descritiva, salienta um fato característico quando se trata de pesquisas no ramo da psicologia. Por vezes, não é possível medir uma variável da maneira mais fidedigna através de escalas de avaliação ou outros instrumentos, pois entre seus extremos há graus intermediários, que tais recursos não abrangem, como por exemplo: avaliar o êxito em determinada profissão. E é quando situações como esta se apresentam é que há a inter-relação da estatística descritiva com a utilização das variáveis nominais, que transforma uma variável em atributo, o que permite a realização da análise, através do cruzamento de informações e disposição das mesmas em tabelas.

Um modo de apresentar relações entre dados qualitativos categóricos bastante utilizado na estatística descritiva é a Tabela de Contingência, cujas linhas e colunas correspondem às variáveis analisadas. Tem por objetivo sintetizar os dados das mesmas, de duas ou mais variáveis, a fim de cruzá-los, extrair as informações possíveis e analisar as relações entre elas, ou seja, verificar se há dependência ou associação. Segue abaixo exemplo de tabela de contingência 2x2 (Jelihovschi, 2014).

Exemplo 1

No ano de 2014, realizou-se uma pesquisa com estudantes de três escolas, uma privada e duas públicas, na cidade de Uberlândia. A pesquisa baseou-se no individualismo vertical, que se caracteriza pelas relações de troca e atitudes do sujeito que está em busca de destaque e

apresenta objetivos independentes dos grupais, e considera que adquirir status social e se esforçar para vencer todo tipo de competição é essencial. Pessoas que seguem essa orientação, valorizam a exaltação dos ganhos e tendem a considerar que a desigualdade faz parte do processo natural da concorrência. (Fleury, 2014; Triandis e Gelfand, 1998; Triandis & Suh, 2002).

O questionário foi aplicado em sala de aula em estudantes do segundo ano do ensino médio, com média de idade de 16 anos. Em termos de informações sociodemográficas preencher as seguintes informações: sexo, cor e escolaridade da mãe. Aqui interessa conhecer suas respostas à seguinte sentença, que é um item traduzido livremente da escala de individualismo vertical de Triandis e Gelfand (1998): “**vencer é tudo para mim**”. Deveriam escolher uma opção dentre: “discordo”, “discordo mas não muito”, “concordo” ou “concordo mas não muito”. A tabela de contingência cruzando as respostas e a variável sexo é apresentada na Tabela 1.

Tabela 1: Tabela de contingência cruzando sexo dos participantes (n = 675) e respostas ao item “Vencer é tudo para mim”

Sexo	Vencer é tudo para mim				Total
	Discordo	Discordo mas não muito	Concordo, mas não muito	Concordo	
Feminino	73 (21,4%)	66 (19,4%)	96 (28,2%)	106 (31,1%)	341(100%)
Masculino	69 (20,7%)	54 (16,2%)	119 (35,6%)	92 (27,5%)	334 (100%)
Total	142 (21,1%)	120 (17,8%)	215 (31,9%)	198 (29,3%)	675 (100%)

Pode-se concluir que dos estudantes do sexo masculino, 35,6%, concordam, mas não muito com a afirmação “vencer é tudo pra mim”, e essa foi a resposta com maior frequência relativa, enquanto que entre as garotas, a maioria concorda totalmente com a mesma sentença, 31,08%. E que dentre os que discordam, as estudantes do sexo feminino superam os estudantes

do sexo masculino.

Observando as proporções gerais da amostra (linha Total) e comparando com as proporções associadas ao sexo dos participantes, percebe-se que as respostas são próximas. Há uma tendência maior de os meninos concordarem parcialmente e as meninas totalmente, e estas também tem proporcionalmente mais discordâncias parciais.

Como relatado anteriormente, a tabela é utilizada para facilitar a visualização das informações e conseqüentemente otimizar as análises e verificações a serem realizadas. Um teste bastante utilizado para essa verificação, da correlação das variáveis da tabela de dupla entrada (2×2), é o Qui Quadrado (χ^2).

De acordo com Kellner & Nick (1977) qui quadrado é um teste não paramétrico, que visa encontrar um valor de dispersão entre as variáveis independentes. O objetivo é comparar as possíveis divergências entre as frequências obtidas e as esperadas, o que foi realizado intuitivamente na interpretação da tabela ao se comparar as proporções gerais com as de cada sexo.

Conti (2009) traz que, para medir essa discrepância utiliza-se a fórmula $\chi^2 = \sum [(o - e)^2 / e]$:

χ^2 = qui quadrado

O = frequência observada

e = frequência esperada

$o - e = d$ (desvio)

Considerando que $(o - e)$ seja o desvio padrão, a fórmula também pode se dar da seguinte maneira: $\chi^2 = \sum (d^2 / e)$.

O mesmo autor salienta que quando as frequências observadas se aproximam das esperadas, o valor de χ^2 é pequeno. Entretanto, quando acontece o inverso, χ^2 assume altos valores, podendo então haver duas hipóteses: nula ou alternativa. Na primeira, significa que não

há associação entre as variáveis. Enquanto que na segunda a associação é mais provável, já que existem diferenças entre as frequências observadas e esperadas.

Rumsey (2012) acrescenta que mesmo que o objetivo das tabelas sejam facilitar a compilação de dados, nem sempre as mesmas são fáceis de ler, pois dependendo da quantidade de informações obtidas, elas se tornam complexas durante a visualização, o que dificultaria a compreensão dos dados, como pode ser visto no segundo exemplo abaixo.

Exemplo 2

Os dados desse segundo exemplo foram retirados da mesma pesquisa citada no exemplo 1. Porém foram utilizadas quatro variáveis: cor, escolaridade, opção de resposta e sexo, apresentadas na Tabela 2.

Tabela 2: Tabela de contingência cruzando sexo dos participantes (n = 675), escolaridade da mãe e cor da pele e respostas ao item “Vencer é tudo para mim”

		Discordo	Discordo, mas não muito	Concordo, mas não muito	Concordo	Total	
Branco	Feminino	Ens. Méd.	22 (25,3%)	13 (14,9%)	19 (21,8%)	33 (37,9%)	87 (100%)
		Ens. Sup.	26 (29,5%)	24 (27,2%)	30 (34,1%)	8 (9,1%)	88 (100%)
	Masculino	Ens. Méd.	10 (14,9%)	11 (16,4%)	25 (37,3%)	21 (31,3%)	67 (100%)
		Ens. Sup.	6 (10,3%)	17 (29,3%)	21 (36,2%)	14 (24,1%)	58 (100%)
Negro	Feminino	Ens. Méd.	17 (15,3%)	15 (13,5%)	30 (27,1%)	49 (44,1%)	111 (100%)
		Ens. Sup.	29 (43,3%)	13 (19,4%)	15 (22,4%)	10 (14,9%)	67 (100%)
	Masculino	Ens. Méd.	17 (13,2%)	13 (10,1%)	54 (41,9%)	45 (34,9%)	129 (100%)
		Ens. Sup.	10 (20,8%)	11 (22,9%)	16 (33,3%)	11 (22,9%)	48 (100%)
TOTAL		137 (20,9%)	117 (17,8%)	210 (32,1%)	191 (29,2%)	655 (100%)	

A Tabela 2 permite identificar que uma alta proporção de mulheres negras com mães com nível superior que discordam do item, e um contraste entre mulheres e homens em que estas concordam absolutamente em maior proporção que estes com a frase na escolaridade nível médio, para brancos e negros. No nível superior, o padrão se inverte; são os meninos que têm maiores proporções de concordância absoluta, para brancos e negros. Porém, é uma tabela de difícil visualização. Optou-se por calcular percentuais para a distribuição de respostas para cada combinação das variáveis sociais, mas poderia haver outras segmentações relevantes que não são facilmente observáveis com essa disposição. Através dos dois exemplos dados acima, observou-se que no primeiro, utilizando a tabela de contingência 2×2 , a observação e

cruzamento dos dados se deu de maneira, simples, direta e objetiva. Em contrapartida, quando se incluíram mais informações na mesma (Tabela 2), abrangendo quatro variáveis diferentes, em um primeiro momento visual, a tabela se tornou mais carregada, exigindo uma concentração maior para encontrar padrões importantes entre os dados apresentados.

O objetivo deste trabalho é contribuir para a solução da complexidade desse segundo caso mencionado, as tabelas de contingências com mais de duas variáveis. O intuito é apresentar uma técnica gráfica de leitura dessas tabelas, os gráficos mosaicos, bem como explicar que eles podem facilitar as análises dos dados de pesquisas que utilizam muitas variáveis. Finalmente, colaborar para que as informações coletadas numa pesquisa, da área psicossocial por exemplo, não se percam e nem sejam preteridas, no momento da exposição, por falta de dispositivo que comportam a visualização geral de modo compreensível.

Desenvolvimento

Gráfico mosaico é uma técnica gráfica que compila informações e ilustra dados de duas ou mais variáveis de modo proporcional aos resultados obtidos. São retângulos expostos de maneira estratégica, cuja posição, tamanho e a distância entre eles representam os valores da frequência de cada variável, dispostos verticalmente ou horizontalmente, o que ajuda na comparação e análise dos resultados e na identificação de dependência ou independência entre as mesmas. (Hartigan & Kleiner 2015)

Eddy (1981) traz que em mosaicos bivariados o tamanho da área de cada retângulo vai depender do valor da frequência da variável. Se os valores forem iguais, os tamanhos das áreas serão semelhantes, mas se um valor for maior que o outro, as áreas terão tamanhos respectivos. As bases são as mesmas. O que pode variar são as alturas de cada uma caso haja distanciamento da independência. Caso contrário, as alturas permanecem as mesmas. E uma linha na horizontal é tracejada no mesmo nível em todos os retângulos do mosaico, para ilustrar onde eles teriam independência. Dessa maneira, é possível, à primeira vista, ter um panorama geral dos

resultados da coleta. Já nos casos multivariados, o mesmo autor revela que um retângulo é feito para cada variável e posteriormente são subdivididos, para que a proporção de todas as variáveis seja representada em cada um.

De acordo com Hartigan e Kleiner (2015) os retângulos dos mosaicos são moldados de acordo com a necessidade de cada situação. Dependendo da quantidade de variáveis utilizadas, a largura, tamanho, altura, espaçamento dos desenhos, são flexibilizados. Entretanto, os mesmos salientam que até seis seria um número viável de variáveis cruzadas a serem trabalhadas.

Como já sabido, no gráfico mosaico, a ordem das variáveis na tabela é importante e a exibição segue estritamente essa ordem, de modo que selecionar a posição correta dos retângulos é essencial. E uma variação dos mosaicos, que também abrangem essa simplificação dos dados, são os gráficos de dois andares, também chamado de *double-decker plots* (Chen, Unwim, & Härdle, 2008).

Segundo Chen, Unwim e Härdle (2008), os gráficos de dois andares são uma maneira de visualizar associações dentro da estrutura de toda a tabela de contingência. Neste caso, os compartimentos são divididos horizontalmente em um gráfico de dois andares, desenhados lado a lado e contendo a mesma altura. Na parte de baixo do gráfico, é feita uma barra que contém todas as variáveis. As listras mostram diferentes tons de cinza, e cada tom representa uma categoria da variável correspondente. E para que a compreensão seja mais concreta, serão dado alguns exemplos dos gráficos citados acima.

Exemplo 3

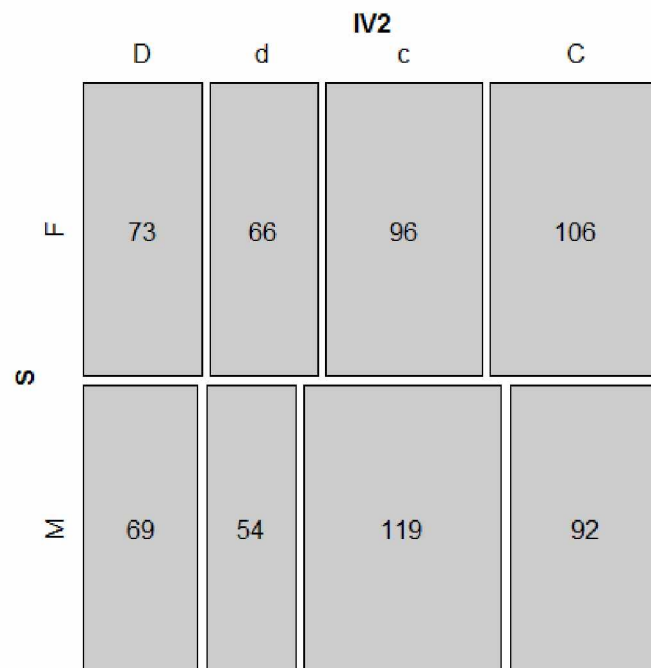


Figura 1: Gráfico mosaico bivariado.

Na Figura 1, vê-se um gráfico mosaico bivariado, utilizando as variáveis: “sexo” (S) e “opção de resposta” (IV2, a variável “Vencer é tudo para mim” dos exemplos 1 e 2, com as opções D, d, c e C); são os dados da Tabela 1. Os retângulos da parte superior são correspondentes as respostas das participantes do sexo feminino, e os inferiores do sexo masculino. Verticalmente, cada um representa a proporção de resposta de acordo com as opções: “discordo” (D), “discordo mas não muito” (d), “concordo” (c), “concordo mas não muito” (C).

Também baseados na sentença “Vencer é tudo pra mim”, nota-se que a maioria dos participantes do sexo masculino concordam com a afirmação, representado pelo retângulo de maior expressão, enquanto que entre as meninas a diferença proporcional entre a concordância parcial e absoluta é menor.

Exemplo 4

Na Figura 2, temos um gráfico mosaico multivariado, abordando as seguintes variáveis:

“sexo” (S), “cor” (C), “opção de resposta” (IV2) e “escolaridade” (E) das mães dos estudantes; são os dados da Tabela 2.

Observe que na lateral esquerda do gráfico é possível identificar uma divisão horizontal em dois blocos. Esses blocos representam a variável “sexo”. O superior, “feminino” (F) e o inferior, “masculino” (M). Na lateral direita, esses dois blocos maiores foram subdivididos horizontalmente pela variável “cor”, de modo que cada bloco foi separado entre “brancos” (B) e “negros” (N). Ainda no sentido horizontal do gráfico nota-se, do lado esquerdo, que há uma segunda separação, “nível superior” (S) e “ensino médio” (B) dos pais dos jovens participantes da pesquisa. Por fim, verticalmente, temos as proporções de acordo com as opções de resposta: “discordo” (D), “discordo mas não muito” (d), “concordo” (c), “concordo mas não muito” (C).

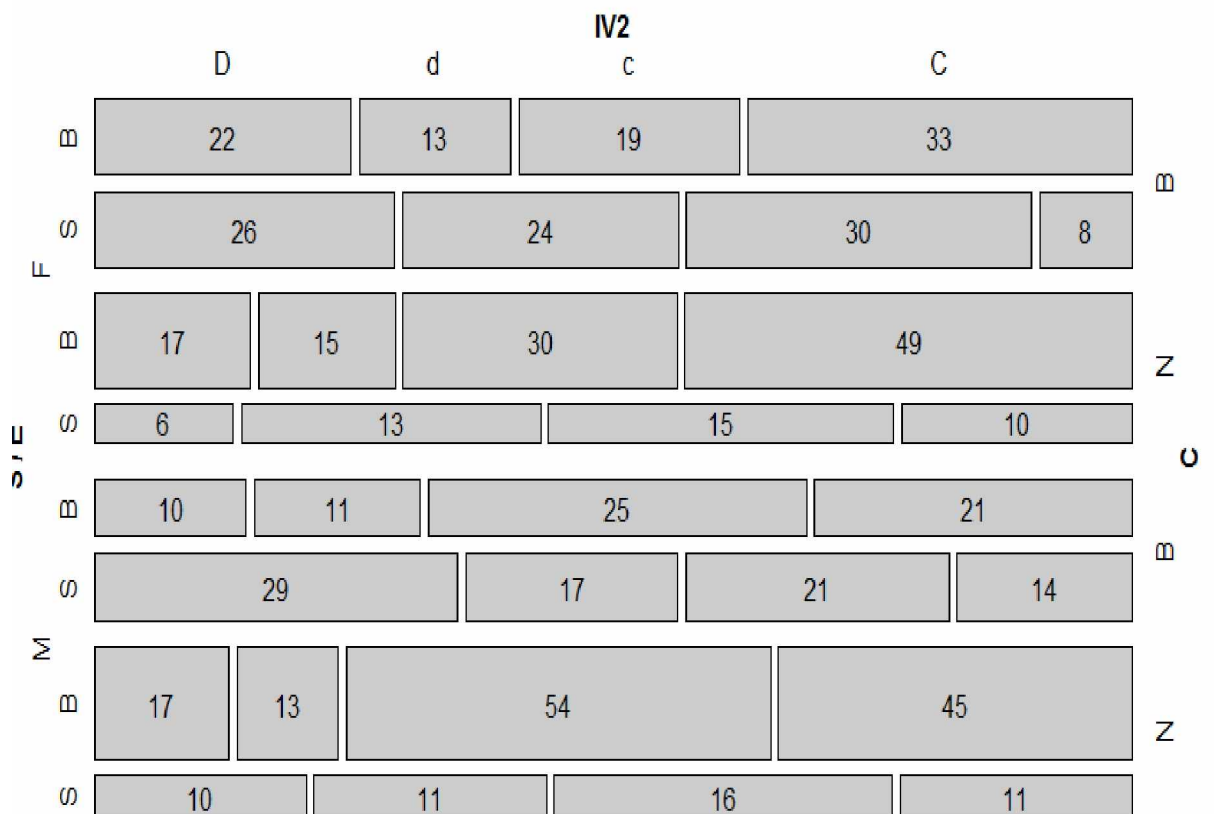


Figura 2: Gráfico mosaico multivariado

Com a demonstração acima, chega-se à conclusão que a maioria dos participantes que

concordam com a sentença são negros, do sexo masculino, cujas mães possuem escolaridade até o ensino médio, pois está representado pelo retângulo mais expressivo. Enquanto que os participantes que mais discordaram da afirmação são brancos, do sexo feminino, com pais que possuem escolaridade de nível superior.

Exemplo 5

Nesta seção traz-se duas figuras de gráficos de dois andares – *double-decker*, a primeira contendo duas variáveis e a segunda quatro. A lógica segue a mesma dos mosaicos padrões, a diferença é a disposição dos retângulos e a posição das variáveis. Contudo, visualmente, consegue-se enxergar a exposição dos dados de maneira mais organizada e sequencial.

A concordância “concordo” e “concordo, mas não muito”, é muito elevada entre negros e negras com mãe com baixa escolaridade. Entre brancos, de ambos os sexo, com mãe com baixa escolaridade também, apesar da tendência de as mulheres concordarem um pouco menos que os homens. Já entre os que tem mães com nível superior, especialmente brancos, as concordâncias são bem menores, e diferenças de sexo também.

No extremo oposto à concordância mencionada acima, “discordo” e “discordo, mas não muito”, observa-se que entre os negros, tanto do sexo masculino quanto do feminino, os que possuem mães com nível de escolaridade superior, estão em menor quantidade, comparados aos que possuem pais com escolaridade de nível médio. Enquanto que entre os brancos, considerando ambos os sexos também, a mesma concordância é elevada independente da escolaridade, exceto entre os homens cujos pais possuem escolaridade nível médio, que apresenta um resultado baixíssimo, comparado aos demais.

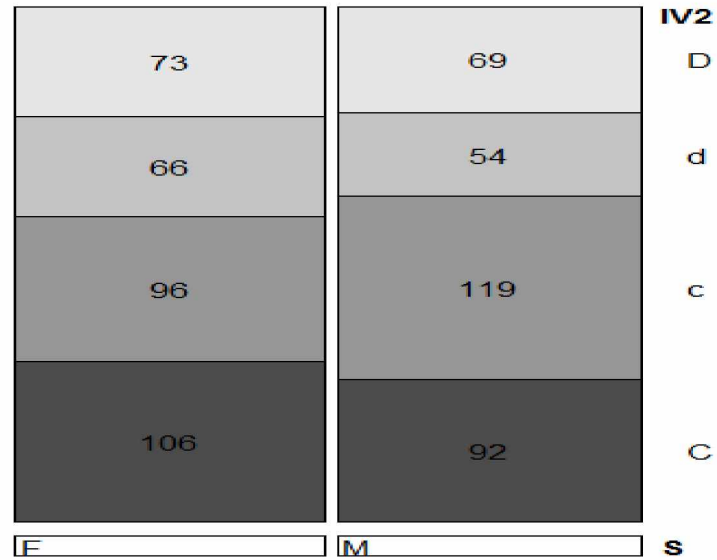


Figura 3: Gráfico de dois andares – double-decker 2 variáveis

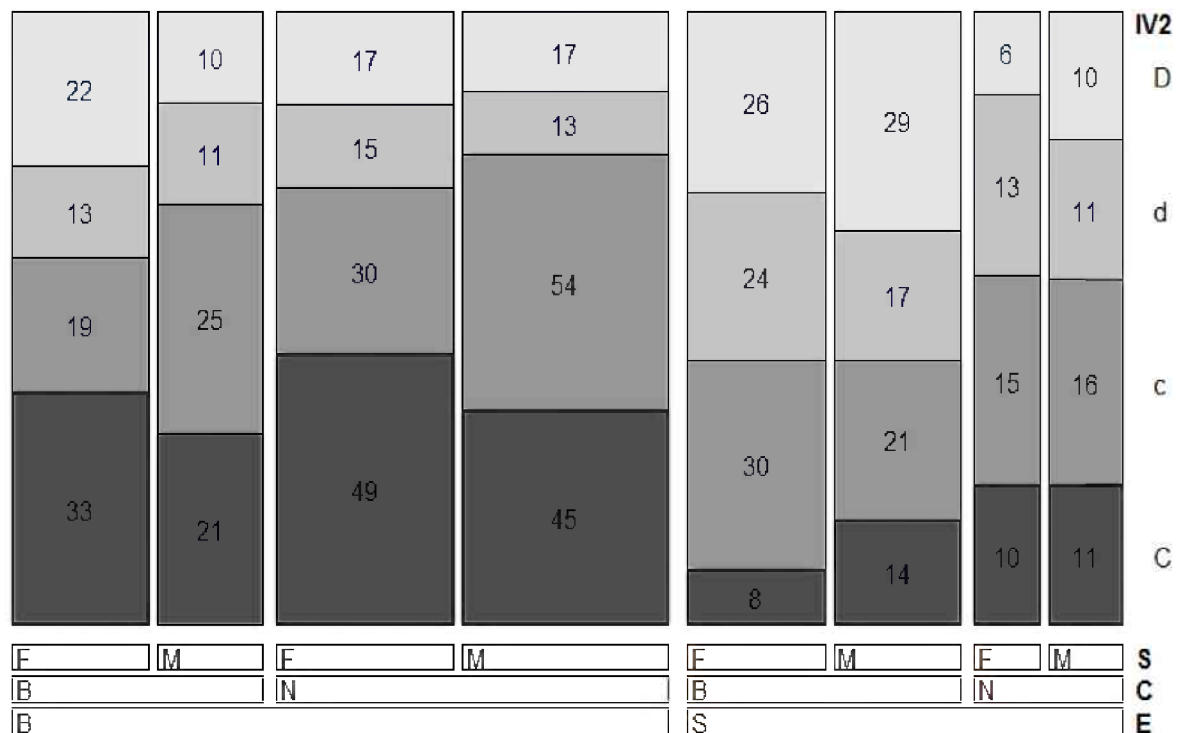


Figura 4: Gráfico de dois andares – double-decker 4 variáveis

A relevância dos mosaicos se dá pelo fato deles colaborarem com o sistema visual humano, transformando informações numéricas em algo ilustrativo. A priori, os mosaicos são

monocromáticos e com o mesmo tracejado, dessa maneira, os desvios de independência existentes, podem ser vistos através do tamanho relativo dos retângulos. Sendo assim, Friendly (2014) relata que para aumentar o impacto visual nesses casos, os resíduos podem ser preenchidos com cores e sombreamentos, indicando desvios com maior ou menor significância.

Já que, em mosaicos monocromáticos a sinalização negativa ou positiva de um desvio de independência não pode ser visualizada, estabeleceu-se um código de cores para essa representação. Células com pequenos resíduos não têm cor. Se o resíduos for médio sua área é sombreada. Em casos de células com valores absolutos menores que 2, seus retângulos não são totalmente cheios e as cores variam entre, azul quando for positivo e vermelho quando for negativo, com tonalidade mais clara. Todavia, quando os valores alcançam escalas maiores, são representados por retângulos completamente preenchidos de azul ou vermelho, com tonalidade escura, quando positivo ou negativo, respectivamente (Friendly, 2014).

Considerações finais

Existem dispositivos que incrementam os gráficos mosaicos, tornando-os mais completos e detalhistas, como o gráfico de peneiras, os gráficos de associação e a utilização de cores em ambos os casos.

De acordo com Chen, Unwim e Härdle (2008), o gráfico de peneiras e os mosaicos são semelhantes estruturalmente. O que os difere é que o primeiro, tem o intuito de exibir a tabela de frequência esperada, o que causa algumas diferenciações em sua formulação. Sendo assim, cada bloco é desenhado proporcionalmente as frequências esperadas das variáveis, e os mesmos são preenchidos com ladrilhos retangulares proporcionais as frequências observadas. São um complemento dos mosaicos, pois detectam estes padrões de dependência.

Em relação a segunda técnica complementar aos mosaicos, os gráficos de associação, Chen, Unwim e Härdle (2008) trazem que devem ser usados apenas quando identificar modelos

de independência seja indispensável, pois representam diretamente o Teste de Person e desvios padronizados e não padronizados. Neste caso cada frequência é representada por um retângulo que tem uma altura proporcional ao residual e largura proporcional à raiz quadrada das frequências esperadas, de modo que a área do mesmo seja proporcional à diferença entre as frequências esperadas e observadas.

A fim de incrementar essas técnicas descritas acima, faz-se o uso de cor e sombreamento, pois ajudam a detectar alguns padrões. Contudo, é necessário ter cuidado na escolha das mesmas, pois, por exemplo: cores de alto croma em áreas grandes, podem ser pontos de distração; cores claras tendem a dar impressão de um espaço maior do que as escuras. Portanto, ao escolher a paleta de cores é necessário certificar-se que uma cor não vai ocupar a posição de mais importante que outra de modo descontrolado. Por isso é importante prestar atenção na saturação, brilho, luz, contraste, em casos de utilizar programas de computador.

Cores uniformes e paletas sempre são utilizadas para exibição das áreas dos quadrados, enquanto que os sombreamentos, são usados para adicionar informações aos gráficos básicos e para auxiliar a análise (Chen, Unwim e Härdle 2008).

Dessa maneira, Chen, Unwim e Härdle (2008) conclui que gráficos mosaico, de associação e de peneiramento podem ser usados para visualizar tabelas multivariadas e conseguem convertê-las em representações planas. De modo que o primeiro é útil para análise exploratória, enquanto os demais se especificam em representar modelos de independência.

Em síntese, os gráficos mosaicos, e os de dois andares de que são mera variação, permitem tornar mais intuitiva a avaliação de interações entre variáveis categóricas. Evidentemente, é necessário respeitar limites; a partir de cinco ou seis variáveis começa a tornar-se difícil a identificação de padrões mais complexos. Em psicologia ainda não são muito utilizados, especialmente no contexto brasileiro, o que indica a existência de nicho para

aprofundar análises descritivas e exploratórias e avançar na consideração de interações, comumente negligenciadas.

Referências

- Cabral, J. A. S., & Guimarães, R. C. (1997). *Estatística*. Lisboa, POR: McGraw-Hill.
- Campos, G. *Estatística prática para docentes e pós graduandos*. Disponível em: <http://www.forp.usp.br/restauradora/gmc/gmc_livro/gmc_livro_cap02.html> Acesso em: 17 de setembro de 2018.
- Chen, Härdle, & Unwim (2008). *Handbook of Data Visualization*. Berlin, DE: Springer.
- Conti, F. *Estiatica-Qui quadrado*. Disponível em: <<http://www.ufpa.br/dicas/biome/biopdf/bioqui.pdf>> Acesso em: 15 de setembro de 2018.
- Fleury, L. M. L. (2014). *A influência do individualismo-coletivismo na intenção de compra online e na inovatividade do consumidor* (Monografia). Universidade de Brasília, Brasília. Disponível em: <<http://bdm.unb.br/handle/10483/10690>> Acesso em: 15 de outubro de 2018.
- Friendly, M. (2014). Mosaic Displays for Multi – Way Contingency Tables. *Journal of the American Statistical Association*. 89(425), 190-200.
- Gil, A. C. (2008). *Métodos e Técnicas de Pesquisa social*. São Paulo, SP: Atlas S.A.
- Gonçalves, F. A. (1977). *Introdução à Estatística: Estatística descritiva*. São Paulo, SP: Atlas S.A.
- Hartigan, J. A., & Kleiner, B. (1984). A Mosaic of Television Ratings. *The American Statistician*. 38(1), 32-35.
- Jelihovschi, E. (2014). *Análise exploratória de dados usando o R*. Ilhéus, BA: Editus.
- Kellner, S. R. O., & Nick, E. (1977). *Fundamentos de Estatística para as ciências do comportamento*. Rio de Janeiro, RJ: Renes.
- Mirshawka, V. (1972). *Estatística*. São Paulo, SP: Livraria Nobel S/A.
- Rumsey, D. (2012). *Estatística para leigos*. Rio de Janeiro, RJ: Alta Books.
- Stevenson, W. J. (1981). *Estatística aplicada à Administração*. São Paulo, SP: Harbra.

Tiboni, C. G. R. (2010). *Estatística Básica: para os cursos de Administração, Ciências Contábeis, Tecnológicas e de Gestão*. São Paulo, SP: Atlas S.A.

Triandis, H. C., & Gelfand, M. J. (1998). Converging measurement of horizontal and vertical individualism and collectivism. *Journal of Personality and Social Psychology*, *74*(1), 118–128.

Triandis, H. C., & Suh, E. M. (2002). Cultural influences on personality. *Annual Review in Psychology*, *53*(1), 133–160.