

**UNIVERSIDADE FEDERAL DO RIO GRANDE DO SUL
FACULDADE DE MEDICINA
PROGRAMA DE PÓS-GRADUAÇÃO EM EPIDEMIOLOGIA**



TESE DE DOUTORADO

Influência da suposição sobre a frequência do consumo, diário ou episódico, e do uso de covariáveis na distribuição do consumo alimentar usual nos métodos MSM, NCI e SPADE

Greice Helen da Costa Laureano

Orientador: Prof. Dra. Suzi Alves Camey

Co-orientador: Prof. Dra. Vanessa Bielefeldt Leotti

Porto Alegre, Setembro de 2018.

**UNIVERSIDADE FEDERAL DO RIO GRANDE DO SUL
FACULDADE DE MEDICINA
PROGRAMA DE PÓS-GRADUAÇÃO EM EPIDEMIOLOGIA**

TESE DE DOUTORADO

Influência da suposição sobre a frequência do consumo, diário ou episódico, e do uso de covariáveis na distribuição do consumo alimentar usual nos métodos MSM, NCI e SPADE

Greice Helen da Costa Laureano

Orientador: Prof. Dra. Suzi Alves Camey

Co-orientador: Prof. Dra. Vanessa Bielefeldt Leotti

A apresentação desta tese é exigência do Programa de Pós-graduação em Epidemiologia, da Universidade Federal do Rio Grande do Sul, para obtenção do título de Doutor.

Porto Alegre, Brasil
2018

CIP - Catalogação na Publicação

Laureano, Greice Helen da Costa

Influência da suposição sobre a frequência do consumo, diário ou episódico, e do uso de covariáveis na distribuição do consumo alimentar usual nos métodos MSM, NCI e SPADE / Greice Helen da Costa Laureano. -- 2019.

117 f.

Orientadora: Suzi Alves Camey.

Coorientadora: Vanessa Bielefeldt Leotti.

Tese (Doutorado) -- Universidade Federal do Rio Grande do Sul, Faculdade de Medicina, Programa de Pós-Graduação em Epidemiologia, Porto Alegre, BR-RS, 2019.

1. Estimaco de consumo alimentar usual. 2. Comparaco de mtodos de estimaco. 3. NCI. 4. SPADE. 5. MSM. I. Camey, Suzi Alves, orient. II. Leotti, Vanessa Bielefeldt, coorient. III. Ttulo.

Elaborada pelo Sistema de Gerao Automtica de Ficha Catalogrfica da UFRGS com os dados fornecidos pelo(a) autor(a).

BANCA EXAMINADORA

Prof. Dr. Alvaro Vigo, Programa de Pós-graduação em Epidemiologia, UFRGS.

Prof. Dr. Eliseu Verly Junior, Departamento de Epidemiologia do Instituto de
Medicina Social da UERJ.

Prof. Dra. Sandra Patrícia Crispim, Departamento de Nutrição, UFPR.

Dedico essa tese aos meus pais Sirlei Laureano e Santo Laureano, aos meus amores Rodrigo Coster, Zelda e Appa que me acompanharam com compreensão e carinho nesta jornada.

AGRADECIMENTOS

Agradeço a minha família, aos meus amigos – em especial à Josiene S. Niesciur e ao Ueldo M. P. Rodrigues – e as minhas orientadoras pelo apoio e incentivo em todos esses anos.

SUMÁRIO

| | |
|--|----|
| ABREVIATURAS E SIGLAS | 8 |
| RESUMO | 9 |
| 1. APRESENTAÇÃO | 13 |
| 2. INTRODUÇÃO | 14 |
| 3. REVISÃO DE LITERATURA | 16 |
| 3.1. ESTIMAÇÃO DO CONSUMO ALIMENTAR | 16 |
| 3.1.1. TIPOS DE ERROS | 16 |
| 3.1.2. DIÁRIO VS EPISÓDICO | 20 |
| 3.1.3. USO DE COVARIÁVEIS | 21 |
| 3.1.4. MÉTODOS DE ESTIMAÇÃO DO CONSUMO ALIMENTAR USUAL | 24 |
| 3.1.4.1. NATIONAL CANCER INSTITUTE (NCI) | 28 |
| 3.1.4.2. MULTIPLE SOURCE METHOD (MSM) | 34 |
| 3.1.4.3. STATISTICAL PROGRAM TO ASSESS DIETARY EXPOSURE (SPADE) 39 | |
| 3.2. JUSTIFICATIVA DO PRESENTE TRABALHO | 43 |
| 4. OBJETIVOS | 44 |
| 4.1. OBJETIVO GERAL | 44 |
| 4.2. OBJETIVOS ESPECÍFICOS | 44 |
| 7. CONCLUSÕES E CONSIDERAÇÕES FINAIS | 80 |

ABREVIATURAS E SIGLAS

| | |
|----------|--|
| AGE MODE | <i>Age-dependent dietary assessment model</i> |
| AIC | Critério de informação de Akaike |
| ANOVA | <i>Analysis of Variance</i> |
| BBN | <i>BetaBinomial-Normal Model</i> |
| DVR | Diferença entre os vícios relativos |
| EFCOVAL | <i>European Food Consumption Validation</i> |
| FFQ | <i>Food Frequency Questionnaire</i> |
| GEE | <i>Generalized Estimating Equations</i> |
| ISU | <i>Iowa State University</i> |
| LNN | <i>LogisticNormal - Normal model with correlation</i> |
| LNN0 | <i>LogisticNormal - Normal model without correlation</i> |
| MM | Modelo Misto |
| MMNL | Modelo misto não linear |
| MSM | <i>Multiple Source Method</i> |
| NCI | <i>National Cancer Institute</i> |
| QFA | Questionário de frequência alimentar |
| R24h | Recordatório alimentar de 24 horas |
| SAS | <i>Statistical Analysis System</i> |
| SPADE | <i>Statistical Program to Assess Dietary Exposure</i> |
| VR | Vício relativo |

RESUMO

Este trabalho foi realizado com o intuito de preencher lacunas da literatura de epidemiologia nutricional no que se refere a avaliação do impacto sobre a utilização das suposições de frequência do consumo alimentar, diário ou episódico, e o uso de covariáveis (Questionário de Frequência Alimentar (QFA), sexo e idade) na estimação da distribuição do consumo alimentar usual para populações dos métodos *National Cancer Institute* (NCI), *Multiple Source Method* (MSM) e o *Statistical Program to Assess Dietary Exposure* (SPADE). Para tanto, foi elaborado um artigo que comparou o impacto da escolha da suposição de consumo diário ou episódico, assim como a influência da utilização do QFA na acurácia dos métodos. Ainda, no segundo artigo, foi verificado se o uso das covariáveis sexo e idade influenciaram a acurácia dos métodos. Em ambos artigos o vício relativo foi a medida de acurácia empregada e os dados de consumo usados foram os de uma pesquisa que coletou 20 Recordatórios Alimentares de 24 horas (R24h) de 302 pessoas residentes na cidade do Rio de Janeiro – Brasil.

Os resultados mostraram que:

- Para componentes alimentares de frequência de consumo diária (até 0,1% de R24h iguais a zero) as estimativas mais acuradas foram as com uso da suposição de consumo diário em todos os métodos.
- Em componentes alimentares com frequência de consumo episódica de nível intermediário (entre 20% e 75% de R24h iguais a zero) as estimativas mais acuradas ocorreram com o MSM fazendo o uso da suposição de consumo episódico.
- Já em componentes alimentares com frequência de consumo episódica de nível baixo (no mínimo 76% de consumos iguais a zero nos R24h), as estimativas mais acuradas utilizando a suposição de consumo episódico foram:
 - No método MSM na estimação dos percentis menores ou iguais a 25 e nos métodos SPADE e NCI na estimação da média e os percentis maiores ou iguais a 50.
- Em relação ao uso do QFA no ajuste dos métodos:
 - Discreto aumento de acurácia dos métodos MSM (principalmente para a média e percentis acima de 75) e SPADE (sobretudo para percentis abaixo de 25).
- Quanto a utilização das covariáveis sexo e idade no ajuste dos métodos:
 - Os métodos não foram impactados para componentes alimentares de frequência de consumo diário;
 - Para as frequências de consumo episódica de nível intermediário:
 - O NCI teve resultados, em geral, piores quando comparado aos sem ajuste;
 - Os métodos MSM e o SPADE não foram impactados.
 - Para as frequências de consumo episódica de nível baixo:
 - O NCI foi impactado, em geral, com piora nos resultados;
 - O MSM sofreu impacto, em geral, com melhora nos resultados;

- O SPADE não foi impactado.

Com esses achados, aconselha-se:

- Utilizar em todos os métodos a suposição de frequência de consumo diário para componentes alimentares de frequência de consumo diária.
- Utilizar a suposição de frequência de consumo episódico em componentes de frequência de consumo episódico de nível intermediário:
 - Emprego do método MSM na estimação dos percentis 5 e 10;
 - Uso do NCI, MSM ou SPADE na estimação dos demais parâmetros da distribuição.
- Utilizar da suposição de frequência de consumo episódico em componentes de frequência de consumo episódico de nível baixo:
 - Emprego do método MSM na estimação dos percentis menores ou iguais a 25;
 - Uso dos métodos SPADE e NCI para estimativa da média e dos percentis acima de 25.
- Utilizar o QFA no SPADE e MSM quando se desejar o máximo de acurácia nas estimações e não houver restrições orçamentárias.
- Fazer o uso de sexo e idade como covariável no método MSM quando se tratarem de componentes alimentares com frequência de consumo episódico de nível baixo.

Palavras-chave: NCI. MSM. SPADE. Consumo episódico. Consumo diário. QFA. R24h.

ABSTRACT

This work was carried out to fill gaps in nutritional epidemiology literature regarding the evaluation of the impact on the use of food frequency assumptions, daily or episodic, and the use of covariates (Food Frequency Questionnaire (FFQ), sex and age) in the estimation of the usual food intake distribution for the methods National Cancer Institute (NCI), Multiple Source Method (MSM) and the Statistical Program to Assess Dietary Exposure (SPADE). To do so, we developed one article with the impact comparison of the choice of the daily or episodic intake assumption, as well as the influence of the QFA use on the accuracy of the methods. On the other hand, in the second article, was verified if the sex and age used as covariable has influenced in the accuracy of the methods. In both articles, the relative bias was the measure of accuracy used and the intake data used data from a survey that collected 20 24-hour dietary recall (24HR) from 302 people living in the city of Rio de Janeiro - Brazil.

The results showed that:

- For daily intake (up to 0.1% 24HR equal to zero) the most accurate estimates was using the daily intake assumption in all methods.
- In food components with intermediate level of episodic intake (between 20% and 75% of 24HR equal to zero), the most accurate estimates occurred with the MSM using the of episodic intake supposition.
- In food components with low level of episodic intake(at least 76% of intakes equal to zero in 24HR) the most accurate estimates using the episodic intake assumption was:
 - MSM method in the estimation of percentiles smaller or equal to 25 and in the SPADE and NCI methods in the estimation of average and percentiles greater or equal to 50.
- About the usage of FFQ:
 - Slight increase in accuracy of MSM (mainly for the average and percentiles above 75) and SPADE (especially for percentiles below 25).
- About the usage of sex and age as covariates:
 - The methods were not impacted for food components of high intake frequency;
 - For intermediate level of episodic intake:
 - The NCI had, in general, worse results when compared to those without adjustment;
 - Both MSM and SPADE were not impacted.
 - For low level of episodic intake:
 - The NCI was generally impacted by worsening results;
 - The MSM had an impact by improving results;
 - SPADE was not impacted.

With these findings, we advised that:

- Use daily assumption in the estimation for food components of high intake frequency, regardless of method;

- Use episodic assumption in intermediate intake frequency components:
 - MSM method in the estimation of the 5th and 10th percentiles;
 - NCI, MSM or SPADE in the estimation of the other parameters of the distribution.
- Use episodic intake frequency assumption in low intake frequency components:
 - MSM in the estimation of percentiles less than or equal to 25;
 - SPADE and NCI to estimate average and percentiles above 25.
- Use the FFQ in SPADE and MSM when you want maximum accuracy in estimates and there are no budget constraints.
- Make use of sex and age as covariate in the MSM method when dealing with food components with low intake frequency.

Keywords: NCI. MSM. SPADE. Episodic assumption. Daily assumption. FFQ. 24HR.

1. APRESENTAÇÃO

Este trabalho consiste na tese de doutorado intitulada “**Influência da suposição sobre a frequência do consumo, diário ou episódico, e do uso de covariáveis na distribuição do consumo alimentar usual nos métodos MSM, NCI e SPADE**”, apresentada ao Programa de Pós-Graduação em Epidemiologia da Universidade Federal do Rio Grande do Sul, em 28 de Setembro de 2018. O trabalho é apresentado em três partes, na ordem que segue:

- I. Introdução, Revisão da Literatura e Objetivos;
- II. Artigo(s);
- III. Conclusões e Considerações Finais.

Documentos de apoio estão dispostos nos apêndices.

2. INTRODUÇÃO

A distribuição do consumo de componentes alimentares tem sido objeto de diversos estudos (1–3). O conhecimento da distribuição é importante para identificar o comportamento alimentar da população em análise e com isso, por exemplo, avaliar se ela tem consumo adequado dos componentes alimentares (4,5).

Nesse contexto, tem-se que para estimar essas distribuições foram desenvolvidos diversos métodos com diferentes tipos de modelagem. Alguns estão implementados em aplicativos computacionais, tais como: STEM.I (6,7) e STEM.II (7); Wallace (8); ISU (9–11); Gay (12); S-Nusser (13); MSM (14,15); SPADE (16–18); NCI (19–22) e AGE MODE (23). Em geral, os métodos diferem de acordo com a frequência de consumo do componente alimentar cuja distribuição pretende-se estimar e com o uso ou não de covariáveis na estimação.

Com essa variedade de métodos disponíveis, há necessidade de pesquisas que os comparem, para que se possa determinar se eles são equivalentes ou quais métodos são mais precisos ou acurados. Vários estudos dessa natureza já foram conduzidos (9,10,13,16,24–26), mas ainda há poucos estudos explorando a influência nos métodos quanto a escolha da suposição de frequência de consumo (diário ou episódico) e a utilização de covariáveis (20,27,16,24,28,29) na estimação da distribuição de consumo alimentar usual. Salienta-se que em nossa experiência empírica de atendimento a pesquisadores verificamos que a questão da escolha do tipo de suposição de frequência de consumo, assim como a utilização de covariáveis, geram dúvidas nos usuários dos métodos, fazendo com que haja dificuldade para o pesquisador fazer essas definições.

Em vista da falta de referências na literatura que mostrem a influência da escolha da suposição de consumo e do uso de covariáveis nos métodos, este trabalho realizou a

comparação da acurácia das estimações para a distribuição de consumo alimentar usual para populações dos métodos *National Cancer Institute* (NCI), *Multiple Source Method* (MSM) e o *Statistical Program to Assess Dietary Exposure* (SPADE). Essas comparações foram realizadas com base em dados de 302 residentes do Rio de Janeiro – Brasil, com 20 Recordatórios Alimentares de 24 horas (R24h) por indivíduo, onde foram mensurados o consumo de alguns componentes alimentares, ao passo que nesses dados foi avaliando o impacto da escolha da suposição de frequência de consumo (diário ou episódico) e da utilização de covariáveis (Questionário de frequência alimentar (QFA), sexo e idade).

3. REVISÃO DE LITERATURA

Nessa revisão de literatura foram apresentados os conceitos básicos para estimação do consumo alimentar, passando pelos tipos de erros inerentes à mensuração do consumo, assim como as técnicas disponíveis para a correção desses erros, além do panorama atual da utilização de covariáveis nos métodos de estimação da distribuição de consumo alimentar e do detalhamento técnico dos métodos NCI, MSM e SPADE.

3.1. ESTIMAÇÃO DO CONSUMO ALIMENTAR

A seguir foram apresentadas algumas definições relacionadas à estimação do consumo alimentar que serão necessárias para o seguimento do trabalho.

3.1.1. TIPOS DE ERROS

A mensuração de dados de consumo alimentar pode gerar dois tipos de erros: o erro sistemático e o erro aleatório (4,30). Todavia, ao se trabalhar com esses dados, é necessário que esses erros sejam minimizados, para que os resultados provenientes não sejam prejudicados.

O *National Research Council (US) Subcommittee on Criteria for Dietary Evaluation* (4) define que o erro aleatório ocorre quando o entrevistado reporta consumos maiores e/ou menores de forma aleatória. Essa variação pode ocorrer nos consumos diários dos sujeito (variação intraindividual) e nos consumos entre os sujeitos (variação interindividual) (30). Já o erro sistemático envolve grupos das populações, por exemplo, quando todos ou quase todos participantes em estudo que têm baixo nível socioeconômico reportam consumo maior de componentes alimentares mais caros (como por exemplo a carne vermelha).

A seguir foi disposto o detalhamento dos tipos de erro, com base no trabalho do *Webinar Series* (31). Na Figura 1 se tem um exemplo do impacto do erro aleatório na

distribuição de consumo alimentar usual, nela é possível observar que um instrumento com erro aleatório não é viciado, mas pode ser impreciso.

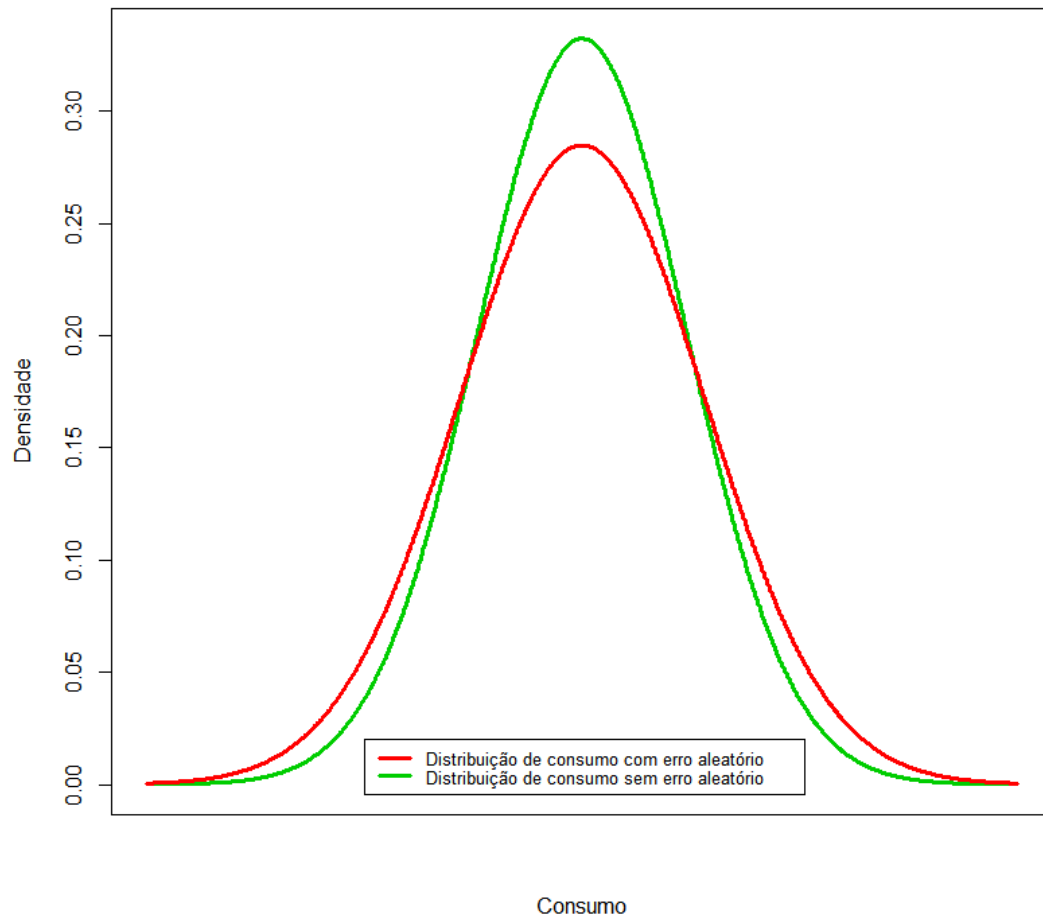


Figura 1: Distribuições de consumo alimentar com e sem erro aleatório.

Na Figura 2 foi disposto o impacto do erro sistemático na distribuição de consumo alimentar usual. Nessa figura é ilustrado que, quando esse tipo de erro está presente, as medições de consumo se afastam do valor verdadeiro na mesma direção, gerando estimativas com viés.

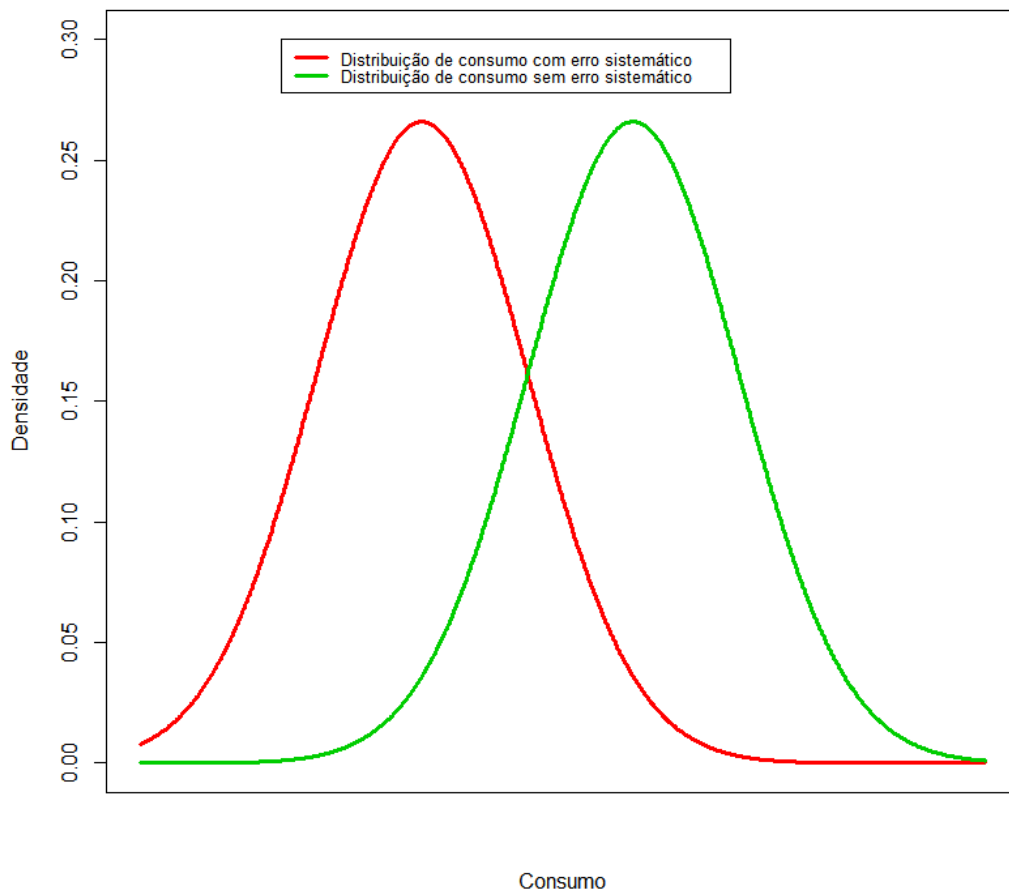


Figura 2: Distribuições de consumo alimentar usual com e sem erro sistemático.

Quando se trata de erros aleatórios, tem-se que a média de várias medidas repetidas aproxima-se do verdadeiro valor do consumo alimentar usual, enquanto que para os erros sistemáticos essa relação assintótica não ocorre (4,5). Com isso, para corrigir o erro sistemático é necessário a utilização de um instrumento de referência que permita estimar a distribuição verdadeira e, com isso, corrigir esse viés. Já o erro aleatório pode ser minimizado pela média de múltiplas medidas repetidas de consumo.

Quando se trata da mensuração do consumo alimentar os instrumentos de coleta mais utilizados são o R24h e o QFA (5). No QFA os indivíduos respondem a respeito da frequência (e, às vezes, da quantidade) em que consomem determinados alimentos em um período. Já no R24h, o indivíduo relata a frequência e a quantidade de tudo que foi

consumido durante as últimas 24 horas.

Apesar do QFA requerer investimento financeiro inferior ao R24h, ele é limitado a uma lista finita de alimentos e é prejudicado pela dificuldade que os indivíduos têm para relatar seu consumo alimentar de um longo período. Em contraste, o R24h fornece detalhes ricos sobre os tipos e quantidades de alimentos consumidos.

Ambas as metodologias têm erros de mensuração atribuídos a elas, sendo que, no caso do QFA os erros sistemáticos são maiores e os aleatórios menores, já no R24h os erros sistemáticos são menores e os aleatórios são maiores (5,31). Apesar dos R24h fornecerem estimativas com menor erro sistemático, a medição de um único dia de R24h é insuficiente para se estimar o consumo de longo prazo, uma vez que geralmente existe grande variabilidade do consumo individual nas populações (5).

A estrutura do erro de medição para o consumo alimentar foi definida a seguir

$$R_{ij} = \beta_0 + \beta_1 T_i + u_i + e_{ij} \quad (1)$$

onde R_{ij} é o consumo reportado do indivíduo i no dia j , β_0 é o erro constante aditivo, β_1 é o erro relacionado ao consumo, T_i é o consumo verdadeiro do indivíduo i , u_i erro relacionado ao indivíduo i e e_{ij} é o erro aleatório.

Dos erros contidos nessa equação, um é referente ao erro aleatório e três são referentes aos erros sistemáticos. Dentre esses erros sistemáticos, tem-se:

- Erro relacionado ao consumo: referente a um valor que é multiplicado pelo consumo verdadeiro dos participantes.
- Erro constante aditivo: erro sistemático que ocorre igualmente para todos os participantes.
- Erro relacionado ao indivíduo: erro sistemático que surge como resultado de características pessoais, como sexo, idade ou peso corporal que levam os sujeitos a relatarem consumos maiores ou menores do que realmente foi ingerido.

Para fazer a correção do erro sistemático tem se utilizado resultados de biomarcadores para se obter a distribuição de consumo verdadeira, e, por conseguinte, remover esse erro. Entretanto, essa solução está disponível apenas para alguns componentes alimentares (5,31).

Com base na estrutura de erros de medição o R24h é a melhor aproximação existente de um instrumento de referência para a aferição do consumo, exceto nos casos para os quais se tem um biomarcador para a atenuação dos erros sistemáticos. Além de haver erro sistemático menor, também é possível corrigir o erro aleatório por meio de modelos estatísticos quando há múltiplas medidas dos R24h por sujeito (5). Salienta-se que além da utilização dos R24h também é possível utilizar o registro alimentar como instrumento de medição de consumo (30).

Por fim, evidencia-se que os erros de medição podem afetar de forma distinta as estimações a nível individual (quando se deseja estimar o valor de consumo de um indivíduo) e a nível populacional (quando se deseja estimar a distribuição de consumo de uma população com base em uma amostra de indivíduos) fazendo com que, por exemplo, sejam necessários um número maior de R24h individuais para estimar a distribuição de consumo individual, quando comparado ao número de R24h individuais para estimar o consumo populacional (30).

3.1.2. DIÁRIO VS EPISÓDICO

Hoffman et al. (13) define o consumo diário quando os componentes alimentares são consumidos diariamente ou quase diariamente por todos indivíduos, por outro lado o consumo episódico é definido quando os componentes alimentares são consumidos raramente ou apenas por uma parte da população. Já no Webinar Series (31) o consumo diário é definido pelo consumo quase diário por quase todas as pessoas e o consumo episódico quando o componente alimentar é consumido episodicamente pela maioria das

peessoas. Em Souverein et al. (16) o consumo diário é definido quando os componentes alimentares são consumidos frequentemente (quase todos os dias) e os de consumo episódico não são consumidos todos os dias.

Apesar de haver as definições já citadas, apenas duas recomendações existentes na literatura determinam de forma quantitativa o consumo diário ou episódico. Em Guenther et al. (32) o consumo diário foi definido quando menos de 10% dos R24h têm consumos iguais a zero. Já no site do NCI (33), tem-se que quando menos de 5% dos R24h têm consumos iguais a zero considera-se consumo diário e consumo episódico quando este percentual for de 10% ou mais. Quando se tratar de percentuais entre 5% a 10% de consumos iguais a zero, a recomendação é ajustar o modelo considerando ora consumo diário e ora consumo episódico e decidir como mais adequado o que obtiver o melhor ajuste aos dados.

3.1.3. USO DE COVARIÁVEIS

Ao se tratar de consumo alimentar, é intuitivo assumir que há diferença no comportamento do consumo entre grupos populacionais, motivo pelo qual as pesquisas que avaliam esta questão, geralmente, mostram os resultados com estratificações destes grupos (3,34–36).

Em Rossum et al. (3), por exemplo, foi avaliada a dieta de crianças e adultos com idades entre 7 e 69 anos através da investigação do consumo alimentar dos holandeses de 2007 a 2010. Nele, os dados dos R24h foram analisados com base em estratificações por sexo e por idade, onde se observou que, para a faixa etária de 19 a 30 anos, o consumo de frutas, nozes e azeitonas de homens ocorreu em 54% dos dias apreciados, à medida que o consumo mediano desses homens foi de 51 gramas/dia, enquanto as mulheres tiveram consumo mediano de 75 gramas/dia em 60% dos dias. Já ao se tratar da faixa etária de 51 a 69 anos, os homens apresentaram consumo em 66%

dos dias, com consumo mediano de 102 gramas/dia. Alternativamente, as mulheres obtiveram consumo em 77% dos dias e consumo mediano de 132 gramas/dia. Com isso, ao comparar os sujeitos mais velhos com os mais jovens, verificou-se que houve aumento da quantidade de consumo de 100% para homens e de 78% para mulheres. Por sua vez, a frequência teve um aumento de 12% para homens e de 17% para mulheres. Contudo, quando se compararam os sexos, verificou-se que os homens consomem em torno de 32% a menos destes alimentos por dia em comparação com as mulheres em ambas as faixas etárias, ao mesmo tempo que a frequência de dias de consumo dos homens variou de 6% a 11% a menos que as mulheres. Evidenciando que nessa população existe diferença no consumo alimentar de frutas, nozes e azeitonas não só entre homens e mulheres, mas também entre faixas etárias. Também em Herrick et al. (36) foi utilizada a estratificação por sexo e por idade, já em Crispim et al. (34) há estratificação por sexo e nacionalidade. Outro exemplo é em Yang et al. (35), que mostra a estratificação de sexo, idade, etnia, grau de escolaridade e status de tabagismo.

Quando se trata da estimação da distribuição de consumo alimentar usual, os métodos de estimação, em geral, permitem a utilização de covariáveis, ao passo que esse uso tem o objetivo de mapear as diferenças entre os grupos populacionais e, conseqüentemente, melhorar a estimação (15,16,18,22,36). Apesar de ser possível usar covariável de qualquer natureza, só o impacto da utilização do QFA já foi avaliado em seis estudos. Em Subar et al. (27) foram analisados componentes alimentares consumidos episodicamente por 965 sujeitos que responderam a quatro R24h e um QFA no estudo *Eating at America's Table* (37). Nos resultados foi observado a existência de correlação positiva entre a probabilidade de consumo (obtida pela estimação da distribuição de consumo alimentar individual do método NCI) e a frequência de dias de consumo relatadas no QFA, assim como houve correlação entre a quantidade média consumida e a

frequência de consumo. Com esses achados, os autores recomendaram a utilização dos dados do QFA como covariável na estimação do consumo usual com o objetivo de reduzir erros.

Adicionalmente, Tooze et al. (20) utilizaram simulação de dois R24h por sujeito, com base nas estimativas do total de grãos do estudo *Eating at America's Table* (37), enquanto que a verdadeira distribuição de consumo foi aproximada com base na média de 365 dias de consumo simulado. Com base nessa verdadeira distribuição, foram comparados dois tipos de suposição do método NCI (consumo episódico – com e sem correlação entre quantidade e frequência de consumo) e a utilização do QFA, chegando-se à conclusão de que o uso do QFA não causou alterações no resultado dos métodos, mas que a suposição de consumo episódico com correlação teve melhores resultados que a sem correlação.

Souverein et al. (16) utilizaram os dados de 293 homens que responderam a dois R24h no estudo EFCOVAL (34), com o objetivo de examinar a influência da inserção do QFA como covariável na qualidade dos métodos NCI e MSM na estimação do consumo de vegetais, frutas e peixe. Nos resultados foi constatado que a inserção do QFA não alterou de forma contundente os métodos.

Por outro lado, em Goedhart et al. (24) utilizaram simulação de dados para analisar o impacto do emprego das informações do QFA na estimação da distribuição de consumo usual. Foi avaliado o impacto do emprego do QFA como covariável e da informação de pessoas que nunca consomem na estimação dos métodos *BetaBinomial-Normal model* (BBN), *LogisticNormal-Normal model without correlation* (LNN0), *LogisticNormal-Normal model with correlation* (LNN e NCI), SPADE e MSM. Chegou-se à conclusão que a utilização do QFA como covariável não influenciou os métodos, já

a informação de quem nunca consome traz benefícios na estimação de percentis mais baixos da distribuição nos métodos.

Em Verly et al. (28) foi utilizado o critério de informação de Akaike (AIC, do inglês *Akaike information criterion*) como medida de qualidade do ajuste dos métodos NCI e MSM, para avaliar a influência de dois tipos de covariável: QFA e uma variável binária indicadora de consumo (ela recebeu valor igual a zero quando não houve consumo em nenhum dos R24h e a frequência de consumo no QFA foi igual a zero; ou recebeu o valor igual a 1 quando houve consumo positivo em qualquer dia dos R24h e/ou a frequência de consumo no QFA não foi nula). Com isso, chegaram à conclusão de que o emprego do QFA como covariável não melhorou a qualidade do ajuste dos modelos. Todavia, a variável indicadora trouxe melhora na qualidade do ajuste dos métodos (principalmente para componentes de consumo episódico).

Em Ost et al. (29) foi avaliado se há ganho em se utilizar as informações do QFA para identificar quem consome e quem nunca consome o componente alimentar na estimação da distribuição de consumo alimentar usual. Para isso, o método SPADE foi ajustado aos dados simulados de três componentes alimentares com proporções diferentes de não consumidores. Os resultados mostram que a indicação dos sujeitos que não consomem não melhorou as estimativas dos percentis mais altos e que, em contrapartida, houve melhora na estimação dos mais baixos.

3.1.4. MÉTODOS DE ESTIMAÇÃO DO CONSUMO ALIMENTAR USUAL

No trabalho de Dodd et al. (5) foi realizada uma revisão geral a respeito da estimação de consumo alimentar usual. Nesse trabalho foi relatado que o consumo usual não é bem retratado com a aferição de um único dia de consumo, e, por isso, começou-se a se utilizar mais de uma medição de R24h por sujeito, ao passo que a estimação da distribuição de consumo usual passou a ser feita pela distribuição empírica das médias

individuais de consumo. Contudo, para muitos componentes alimentares, a média dos R24h ainda contém considerável variação intraindividual, fazendo com que a distribuição das médias individuais possua variância maior do que a verdadeira distribuição de consumo usual, o que leva a uma estimativa imprecisa da distribuição.

Esses potenciais vieses foram ilustrados na Figura 3 que mostra as distribuições de consumo estimadas de consumos gerados aleatoriamente e que representam um R24h por sujeito (linha tracejada), a média de dois R24h por indivíduo (linha tracejada e pontilhada) e a distribuição de consumo usual estimada (linha sólida). Já a linha vertical sólida é referente a média de dois R24h por indivíduo e a vertical tracejada é referente a média das outras duas distribuições.

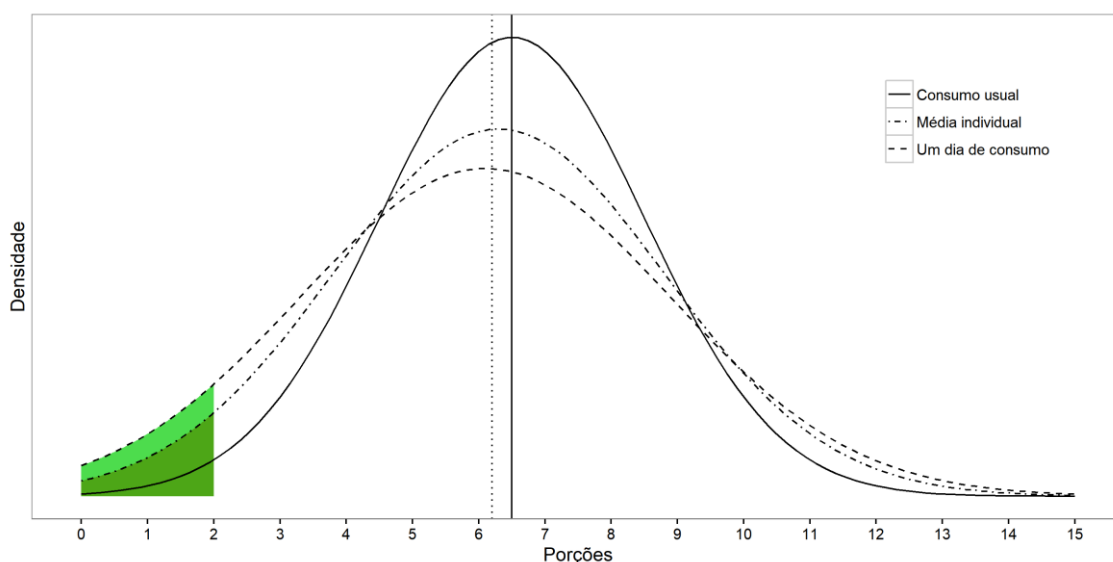


Figura 3: Distribuições de consumo estimadas.

Já as áreas sob as curvas à esquerda das retas verticais são referentes a porcentagem da população com consumo de menos de sete porções por dia (que é aproximadamente 50% para todas as três distribuições). No entanto, quando se avalia se o consumo alimentar está adequado às recomendações, muitas vezes esses valores estão localizados nas caudas da distribuição de consumo, onde os vieses podem ser maiores. Por exemplo, a porcentagem da população com consumo usual de menos de duas porções

por dia é estimada em 5,7%, 3,7% e 1,1%, respectivamente, para um R24h por entrevistado, para a média de dois R24h por indivíduo e para a distribuição de consumo usual estimada. Com base em um exemplo similar ao da Figura 3, Dood et al. (5) relataram que a utilização de uma observação de R24h ou da média de múltiplos R24h é suficiente para estimar o consumo usual médio. Entretanto, são insuficientes para estimar os percentis da distribuição de consumo alimentar usual, sendo necessário a utilização de métodos estatísticos que atenuem a variabilidade intraindividual da distribuição de consumo usual.

Em vista disso, nessa sessão foram apresentados alguns métodos para estimação da distribuição do consumo alimentar usual que fazem essa atenuação da variabilidade intraindividual de consumo, conforme o Quadro 1. Tais métodos foram propostos com a finalidade de fazer a estimação para populações em relação ao consumo de componentes alimentares se valendo dos R24h.

Quadro 1: Relação de métodos para a estimação da distribuição do consumo alimentar usual.

| Método | Ano de publicação | Software | Licença | Permite uso de covariáveis? | Modela o consumo episódico? | Permite a modelagem de mais de um item alimentar? | Permite a definição de pontos de corte para valores recomendados de consumo? | Permite planos amostrais complexos? |
|-------------------|--------------------------|--|----------------|------------------------------------|------------------------------------|--|---|--|
| Gay (12) | 1994 | Programa que tenha a ANOVA* e permita simulação. | - | Não | Não | Não | Não | Não |
| ISU (9–11)** | 1996 | Versão SIDE no SAS, Versão C-SIDE e PCSIDE. | Pago | Sim | Não | Não | Não | Sim |
| STEM.I e II (6,7) | 1996 | Programa que tenha a ANOVA* | - | Sim | Sim | Não | Não | Não |
| Wallace (8) | 1999 | Programa <i>Small</i> | - | - | Não | Não | Não | Não |

| Método | Ano de publicação | Software | Licença | Permite uso de covariáveis? | Modela o consumo episódico? | Permite a modelagem de mais de um item alimentar? | Permite a definição de pontos de corte para valores recomendados de consumo? | Permite planos amostrais complexos? |
|---------------|-------------------|----------|---------|-----------------------------|-----------------------------|---|--|-------------------------------------|
| S-Nusser (13) | 2002 | SAS | Pago | Sim | Não | Não | Não | Sim |
| AGE MODE (23) | 2006 | S-PLUS | Pago | Sim | Não | Não | Não | Não |
| NCI (19–22) | 2009 | SAS | Pago | Sim | Sim | Sim | Sim | Sim |
| SPADE (16–18) | 2010 | R | Livre | Sim | Sim | Sim | Sim | Sim |
| MSM (14–16) | 2011 | Web | Livre | Sim | Sim | Não | Não | Não |

*Análise de Variância (ANOVA, do inglês *Analysis of Variance*).

** Primeiro método com software desenvolvido para estimação da distribuição de consumo usual.

Em geral, os dados de consumo alimentar têm distribuições assimétricas, o que impossibilita a utilização direta de métodos que pressuponham normalidade para estimar a distribuição de consumo de componentes alimentares (5). Para contornar esse problema a maioria dos métodos utiliza a transformação dos dados dos R24h para fazer a melhor aproximação possível da distribuição normal. Esses métodos de estimação da distribuição de consumo usual geralmente têm dois tipos de suposição sobre a frequência de consumo: diário e episódico.

Quando se trata da estimação com suposição de consumo diário, as metodologias utilizam apenas a informação da quantidade consumida pelos indivíduos nos R24h e, após transformar esses dados para a escala normal, realizam a estimação da média e das variâncias interindividual e intraindividual. Posteriormente, faz-se a atenuação do efeito do erro intraindividual por meio da remoção da variância intrarindividual. Assim, após a obtenção das estimativas atenuadas na escala normal, é feita a transformação inversa dos

resultados para a escala original e, então, se obtém a distribuição de consumo alimentar usual.

Já quando se trata da suposição de consumo episódico, os métodos fazem uso da informação da quantidade e da frequência de consumo. Em vista disso, além deles realizarem a estimação considerando a quantidade de consumo (da mesma forma que no consumo usual), eles também estimam a probabilidade de consumo com base na informação de frequência. Após a estimação da quantidade e da probabilidade de consumo, faz-se a multiplicação das estimativas e se obtém a distribuição de consumo alimentar usual.

Os métodos se diferenciam em relação à forma de estimação dos parâmetros (média e variâncias) e da transformação dos dados, mas – em geral – eles têm as seguintes pressuposições (4-12,16,18-22,32):

- O R24h fornece mensurações não viciadas para o consumo usual;
- Os dados transformados podem ser decompostos em dois componentes independentes (variação interindividual e intraindividual);
- A variância intraindividual não influencia no consumo em longo prazo.

Neste trabalho, foram abordados três métodos que permitem escolher se o consumo alimentar é diário ou episódico: NCI, MSM e SPADE. O detalhamento de cada um deles foram dispostos logo a seguir.

3.1.4.1. NATIONAL CANCER INSTITUTE (NCI)

O NCI (20–22) foi proposto para estimar a distribuição do consumo alimentar diário ou episódico de componentes alimentares. Ele foi adaptado ao passar dos anos e neste trabalho foi abordada a versão mais atual (Versão 2.1, de 2015) que está implementada em duas macros do SAS (39):

- **MIXTRAN:** utiliza modelos mistos para ajustar os dados de consumo

alimentar, com o intuito de estimar os parâmetros necessários para a segunda macro;

- **DISTRIB:** utiliza o método de simulação Monte Carlo para estimar a distribuição de consumo usual da população (21,22).

Ambas as macros podem ser encontradas no site:

<http://riskfactor.cancer.gov/diet/usualintakes/>.

3.1.4.1.1. Macro MIXTRAN

A estimação dos parâmetros depende da suposição sobre o consumo do componente alimentar, que pode ser:

- episódico, sem correlação entre frequência e quantidade de consumo;
- episódico, com a frequência de consumo correlacionada com a quantidade de consumo;
- diário.

Para qualquer uma dessas situações, nos dados referentes a quantidade de consumo, a macro ajusta o modelo de equações de estimação generalizadas (GEE do inglês *Generalized Estimating Equations*), com o objetivo de estimar os valores iniciais para os parâmetros. Logo após, esses valores são utilizados como valores iniciais para o ajuste do Modelo Misto Não Linear (MMNL), que faz uso da otimização Quasi-Newton na estimação dos parâmetros, à medida que essa otimização utiliza a verossimilhança aproximada pelo método de Quadratura Gaussiana Adaptativa (40). Apesar das questões em comum, cada suposição citada tem suas peculiaridades na estimação dos parâmetros:

- Consumo episódico sem correlação: faz-se a estimação dos parâmetros de forma separada, uma vez que um modelo é utilizado para as informações de quantidade consumida e outro para frequência de consumo;
- Consumo episódico com correlação: faz-se a estimação dos parâmetros em

um modelo que considera as informações de quantidade consumida, de frequência de consumo e correlação entre quantidade e frequência;

- Consumo diário: faz-se a estimação dos parâmetros em apenas um modelo que considera apenas a informação de quantidade consumida.

O detalhamento do processo de estimação do método NCI foi disposto a seguir. Entretanto, como neste trabalho não vai ser abordada a suposição de existência de correlação entre a frequência e a quantidade, esse processo de estimação não será detalhado.

Suposição de consumo episódico

Para a suposição de consumo episódico sem correlação o método é executado em duas partes: a primeira parte é referente a estimação com base na probabilidade de consumo, e a segunda parte é referente a estimação com base na quantidade de consumo.

Primeira parte:

Na Figura 4 foi disposto um fluxograma que mostra de forma simplificada o processo de estimação dos parâmetros para a probabilidade de consumo.

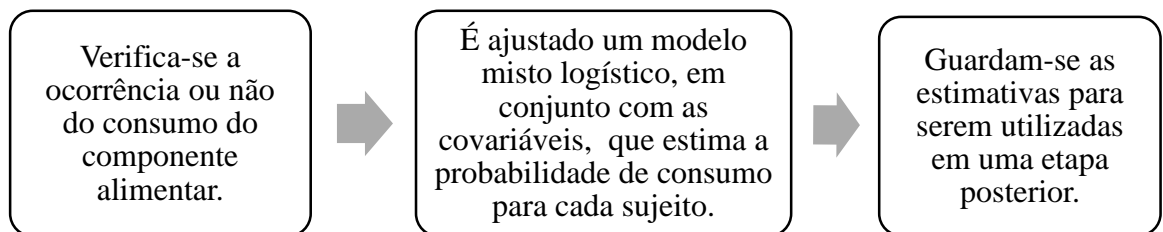


Figura 4: Processo de estimação dos parâmetros para estimar a probabilidade de consumo.

Considerando a variável aleatória Y_{ij} , que representa a quantidade consumida do componente alimentar pelo sujeito i no tempo j . Tem-se que a probabilidade de consumo do componente alimentar é estimada por um modelo misto logístico definido como

$$P(Y_{ij} > 0 | X_{1i}) = p_i = \frac{1}{1 + \exp\left(-(\beta_{10} + \beta_{X1}^t X_{1i} + v_i)\right)} \quad (2)$$

onde p_i é a probabilidade de consumo do indivíduo i , \exp é a função exponencial, $\beta_1 = (\beta_{10}, \beta_{X1}^t)^t$ são os parâmetros de efeito fixo; X_{1i} é o vetor de covariáveis do indivíduo i e v_i é o erro aleatório da probabilidade de consumo que possui distribuição $N(0, \sigma_v^2)$.

Segunda parte:

Na Figura 5 foi disposto um fluxograma que mostra o processo de estimação para a quantidade de consumo.

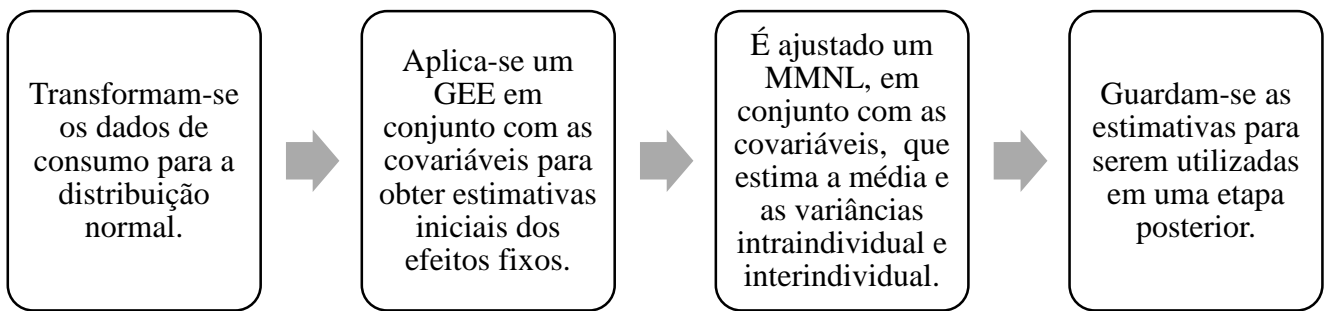


Figura 5: Processo de estimação dos parâmetros para a quantidade de consumo.

Considerando a variável aleatória Y_{ij} , aplica-se a transformação Box-Cox de um parâmetro (41) que é definida como

$$Y_{ij}^* = \begin{cases} \frac{Y_{ij}^\lambda - 1}{\lambda} & \text{se } \lambda \neq 0 \\ \log(Y_{ij}) & \text{se } \lambda = 0 \end{cases} \quad (3)$$

onde Y_{ij}^* são os dados de quantidade de consumo transformados e λ é uma constante que melhor aproxima a distribuição de Y_{ij}^* da distribuição normal (esse valor pode ser estimado pela macro ou fixado arbitrariamente) – salienta-se que não necessariamente os dados transformados terão distribuição normal, uma vez que nem sempre é possível

encontrar o valor de λ que faça essa transformação. Nesses dados transformados é ajustado um GEE e se estima os valores iniciais dos efeitos fixo e aleatório para a quantidade de consumo em função das covariáveis. Assim, tem-se que esses valores são utilizados como valores iniciais no ajuste de um MMNL conforme a equação abaixo:

$$g(Y_{ij}, \lambda) = \beta_{20} + \beta_{X2}^t X_{2i} + u_i + \varepsilon_{ij} \quad (4)$$

onde $g(Y_{ij}, \lambda)$ é a função que transforma os consumos originais para escala normal, $\beta_2 = (\beta_{20}, \beta_{X2}^t)^t$ são os parâmetros de efeito fixo, X_{2i} é o vetor de covariáveis do indivíduo i , u_i é o erro aleatório interindividual que possui distribuição $N(0, \sigma_u^2)$ e ε_{ij} é o erro aleatório intraindividual que possui distribuição $N(0, \sigma_\varepsilon^2)$ - ambos os erros são para a quantidade de consumo.

Salienta-se que os vetores de covariáveis da primeira (X_{1i}) e segunda (X_{2i}) parte podem conter covariáveis distintas.

Suposição de consumo diário

Para a suposição de consumo diário, o método utiliza apenas a segunda parte de modelagem, que é referente a estimação da quantidade de consumo.

Por fim, para ambas as suposições, após a estimação dos parâmetros a macro guarda os valores para posteriormente serem utilizadas pela macro DISTRIB.

3.1.4.1.2. Macro DISTRIB

A macro DISTRIB utiliza simulação de Monte Carlo para simular k consumos, por sujeito, fazendo o uso das estimativas obtidas pela macro MIXTRAN, com o objetivo de estimar a média e os percentis da distribuição do consumo usual. Nessa etapa também há diferença na forma de estimação quando se supõe consumo diário ou episódico.

Suposição de consumo episódico

O processo simplificado de estimação de quando há a suposição de consumo

episódico sem correlação foi apresentado na Figura 6 a seguir

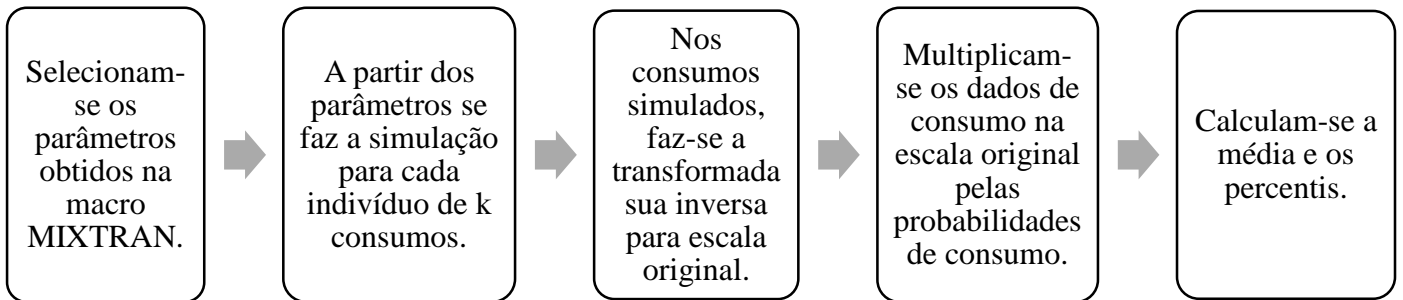


Figura 6: Processo de estimação da distribuição do consumo alimentar usual para quando se supõe consumo episódico.

Primeiramente, é gerado o consumo usual na escala transformada com base na simulação de k consumos por indivíduo. Para isso, primeiramente, com base na variância interindividual (σ_u^2), faz-se a simulação de k realizações do erro aleatório interindividual (u_l), onde $l = 1, \dots, kn$ e n é o número de sujeitos. Posteriormente, os valores simulados de u_l são aplicados na fórmula $\mu_l^* = \beta_{20} + \beta_{X2}X_{2l} + u_l$, onde $X_{2l} = X_{2i}$, e se obtém l consumos simulados na escala transformada.

Após a geração dos consumos simulados, é realizada a transformação inversa para escala original desses consumos. Por fim, multiplicam-se o consumo na escala original e a probabilidade de consumo individual obtida pela equação (2) e, em seguida, calculam-se os percentis e a média dos dados obtidos.

Suposição de consumo diário

Foi disposto na Figura 7 o fluxograma do processo de estimação para a suposição de consumo diário.

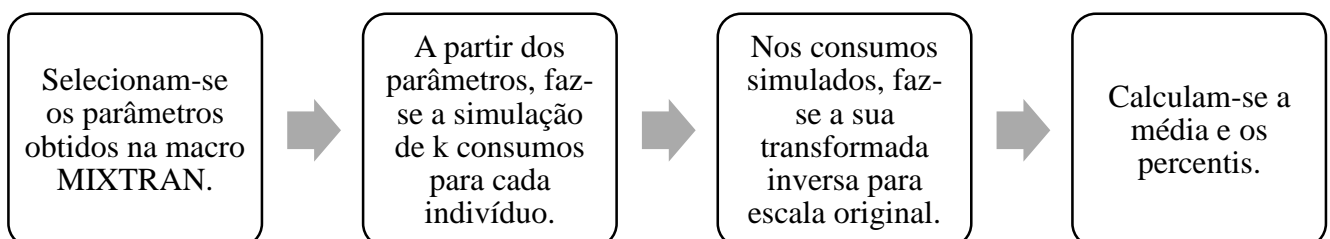


Figura 7: Processo de estimação da distribuição do consumo alimentar usual para quando se supõe consumo diário.

Primeiramente, faz-se a geração dos consumos simulados (conforme descrito na suposição de consumo episódico). Após a criação do banco de dados simulados, eles são transformados para a escala original e se calcula a média e os percentis de forma empírica (42).

3.1.4.2. MULTIPLE SOURCE METHOD (MSM)

O Método MSM foi proposto por Haubrock *et al.* (14) e Harttig *et al.* (15) e está implementado no site: <https://msm.dife.de/>. Para estimar a distribuição de consumo usual esse método permite escolher entre duas suposições sobre o consumo do componente alimentar: diário ou episódico.

Como no NCI, a estimação para cada suposição tem suas peculiaridades, uma vez que na suposição de consumo episódico a estimação é realizada em duas etapas (uma que utiliza as informações de quantidade consumida e outra que usa a probabilidade de consumo) e, na de consumo diário, a estimação é feita com base na informação de quantidade consumida. Os detalhamentos do processo de estimação dos dois tipos de suposição foram mostrados a seguir.

Suposição de consumo episódico

Para a suposição de consumo episódico, o método divide-se em duas partes. Em cada uma é possível utilizar covariáveis, mas – diferentemente do NCI – elas têm que ser as mesmas em ambas as partes da modelagem.

Primeira parte:

Na Figura 8 foi disposto um fluxograma para o processo de estimação da probabilidade de consumo.

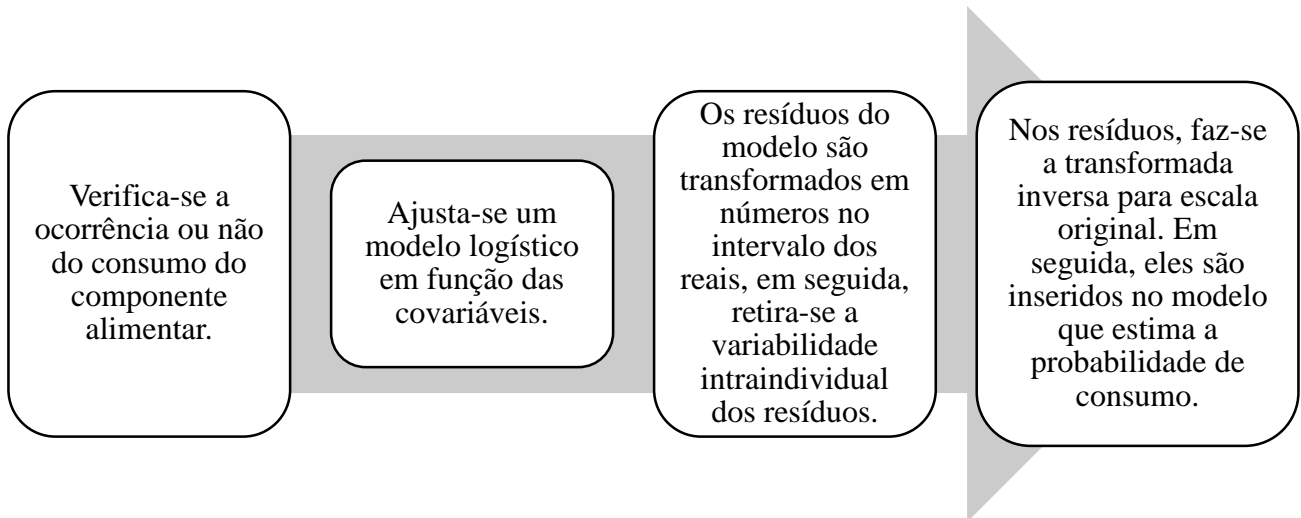


Figura 8: Processo de estimação para a probabilidade de consumo.

Considerando Y_{ij} que representa a quantidade consumida do sujeito i e no dia j , tem-se que p_{ij} , que representa a ocorrência de consumo do sujeito i e no dia j , é definida por

$$p_{ij} = \begin{cases} 0, & \text{se } Y_{ij} = 0 \\ 1, & \text{se } Y_{ij} \neq 0 \end{cases} \quad (5)$$

Com isso, a probabilidade de consumo de um alimento é estimada com um modelo de regressão logística em função do vetor de covariáveis (X_{1i})

$$p_{ij} = \frac{1}{1 + \exp(-(\beta_{10} + \beta_{X1}^t X_{1i} + r_{ij}))} \quad (6)$$

onde \exp é a função exponencial, $\beta_1 = (\beta_{10}, \beta_{X1}^t)^t$ são parâmetros de efeito fixo, X_{1i} é o vetor de covariáveis do indivíduo i e r_{ij} é o resíduo do modelo definido no intervalo $[-1; 1]$

Esse modelo foi definido de forma abreviada pelos autores conforme a equação

$$p_{ij} = m_{i/X_{1i}} + r_{ij} \quad (7)$$

onde $m_{i/X_{1i}}$ é a previsão do modelo logístico para o indivíduo i . Posteriormente, os resíduos do modelo são transformados para números no intervalo dos reais utilizando a função $g(r_{ij}) = \log((1 + r_{ij})/(1 - r_{ij}))$.

Após a transformação dos resíduos é utilizada a decomposição

$$z_{ij} = g(r_{ij}) = t_i + \varepsilon_{ij} \quad (8)$$

onde t_i é uma variável estocástica com esperança E_{t_i} (que é o valor esperado do resíduo do modelo de regressão logística para o indivíduo i), já ε_{ij} é a variação diária do resíduo para o sujeito i . Esses dois termos (t_i e ε_{ij}) são supostos independentes entre si e têm variâncias interindividuais (σ_t^2) e intraindividuais (σ_ε^2), respectivamente.

Para atenuar o efeito da variabilidade intraindividual nos resíduos do modelo de regressão logística o método primeiramente estima a variância interindividual do resíduo

$$\hat{\sigma}_t^2 = \hat{\sigma}_{\bar{z}}^2 - \frac{\hat{\sigma}_\varepsilon^2}{k} \quad (9)$$

onde $\hat{\sigma}_{\bar{z}}^2$ é a variância da média individual do resíduo na escala transformada (\bar{z}_i), $\hat{\sigma}_\varepsilon^2$ é a variância intraindividual estimada do resíduo e k é o número de dias em que os indivíduos foram avaliados.

Posteriormente, faz-se a redução da variância intraindividual nos resíduos do modelo de regressão logística

$$\hat{t}_i = \frac{\hat{\sigma}_t}{\hat{\sigma}_{\bar{z}}} [\bar{z}_i - \bar{z}_{..}] + \bar{z}_{..} \quad (10)$$

onde $\hat{\sigma}_t$ é o desvio padrão interindividual do resíduo, $\hat{\sigma}_{\bar{z}}$ é o desvio padrão da média individual do resíduo e $\bar{z}_{..}$ é a média geral dos resíduos.

Assim, tem-se que a probabilidade de consumo para o sujeito i é definida por

$$p_i^* = m_{i/X_{1i}} + g_{retrans}(\hat{t}_i) \quad (11)$$

onde $g_{retrans}$ é a transformação inversa dos resíduos que é definida por

$$g_{retrans}(\hat{t}_i) = \frac{\exp(\hat{t}_i) - 1}{\exp(\hat{t}_i) + 1} - \frac{\exp(\varpi \bar{t}) - 1}{\exp(\varpi \bar{t}) + 1} \quad (12)$$

onde \bar{t} é a média geral dos resíduos estimados (\hat{t}_i) do modelo de regressão logística e ϖ é raiz quadrada de [1- coeficiente de correlação intra classe].

Segunda parte:

Essa etapa é realizada para os casos de consumo positivo, ou seja, $Y_{ij} > 0$. Antes de detalhar as etapas do método, foi apresentada a Figura 9, na qual foi disposto um fluxograma do processo de estimação.

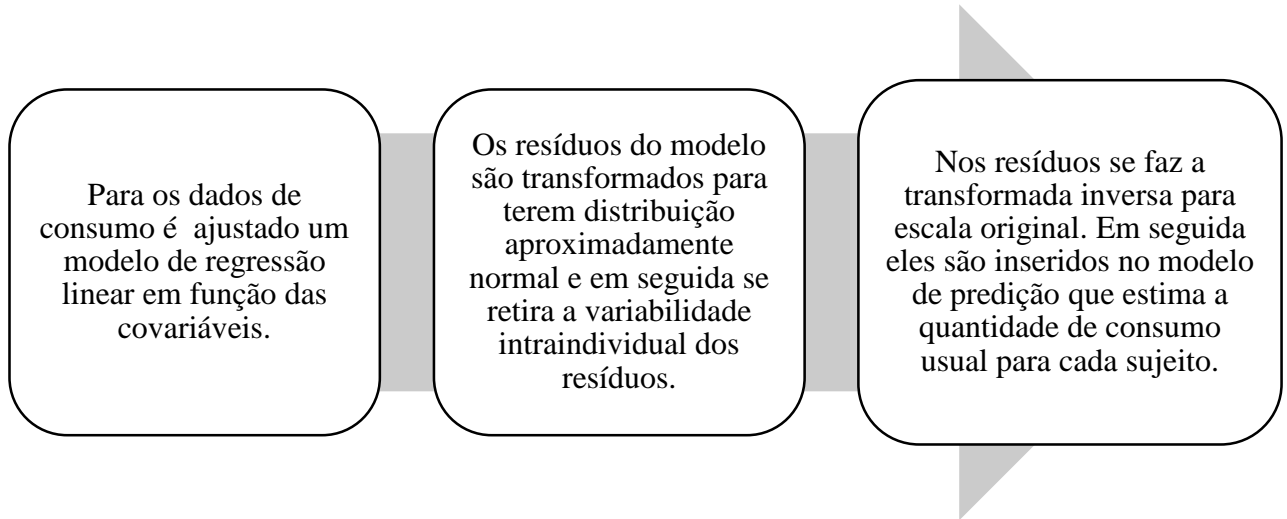


Figura 9: Processo de estimação para a quantidade de consumo.

Nessa etapa os dados de consumo alimentar são ajustados por uma regressão linear em função do conjunto de covariáveis (X_{1i})

$$Y_{ij} = \beta_{20} + \beta_{X1}^t X_{1i} + R_{ij} \quad (13)$$

onde $\beta_2 = (\beta_{20}, \beta_{X1}^t)^t$ são parâmetros de efeito fixo e R_{ij} são os resíduos do modelo.

Esse modelo de regressão foi definido de forma abreviada pelos autores conforme a equação abaixo

$$Y_{ij} = M_{i/X_{1i}} + R_{ij} \quad (14)$$

onde $M_{i/X_{1i}}$ é a previsão do modelo para o consumo do indivíduo i . Logo em seguida, os resíduos do modelo são transformados para ter uma distribuição aproximadamente normal pela transformada de Box-Cox de dois parâmetros (41):

$$f(R_{ij}) = \begin{cases} ((R_{ij} + \omega)^\tau - 1)\tau^{-1} & \text{se } \tau \neq 0 \\ \ln(R_{ij} + \omega) & \text{se } \tau = 0 \end{cases} \quad (15)$$

onde os valores de ω e τ são determinados por maximização de Shapiro-Wilk, onde τ varia ao longo do conjunto de valores $\left[1, \frac{1}{2}, \frac{1}{3}, \frac{1}{4}, \dots, \frac{1}{10}, 0\right]$ e ω varia ao longo do mesmo conjunto multiplicado pela média dos dados originais. Destaca-se que não necessariamente os dados transformados terão distribuição normal, uma vez que nem sempre é possível encontrar os valores de ω e τ que transformem os dados para a distribuição normal.

Após a transformação dos resíduos é utilizada a decomposição

$$Z_{ij} = f(R_{ij}) = T_i + e_{ij} \quad (14)$$

onde T_i é a variável aleatória com esperança E_{t_i} (que é o valor esperado do resíduo do modelo de regressão linear para o indivíduo i) e e_{ij} é a variação diária dos resíduos para o sujeito i . Esses dois termos (T_i e e_{ij}) são supostos independentes entre si com variâncias interindividual (σ_T^2) e intraindividual (σ_e^2), respectivamente.

Para atenuar o efeito da variabilidade intraindividual nos resíduos do modelo de regressão linear o método primeiramente estima a variância interindividual

$$\hat{\sigma}_T^2 = \hat{\sigma}_Z^2 - \frac{\hat{\sigma}_e^2}{k} \quad (16)$$

onde $\hat{\sigma}_T^2$ é a variância interindividual estimada, $\hat{\sigma}_Z^2$ é a variância da média individual do resíduo na escala transformada (\bar{Z}_i), $\hat{\sigma}_e^2$ é a variância intraindividual estimada e k é o número de dias em que os indivíduos consumiram ($k > 0$).

Em seguida, faz-se a redução da variância intraindividual nos resíduos do modelo de regressão linear

$$\hat{T}_i = \frac{\hat{\sigma}_T}{\hat{\sigma}_Z} [\bar{Z}_i - \bar{Z}_{..}] + \bar{Z}_{..} \quad (17)$$

onde $\hat{\sigma}_T$ é o desvio padrão interindividual do resíduo, $\hat{\sigma}_Z$ é o desvio padrão da média individual do resíduo e $\bar{Z}_{..}$ é a média geral dos resíduos.

Por fim, adiciona-se ao modelo de estimação do consumo os resíduos transformados para a escala original.

$$Y_i^* = M_{i/X_{1i}} + f_{retrans}(\hat{T}_i) \quad (18)$$

onde $f_{retrans}$ é a transformação inversa dos resíduos, que é definida conforme a equação a seguir

$$f_{retrans}(\hat{T}_i) = \lambda^{-\lambda} \sum_{s=0}^{int(\lambda/2)} \binom{\lambda}{2s} (\hat{T}_i + \lambda)^{\lambda-2s} \sigma_{\varepsilon}^{2s} (2s-1)!! - \omega \quad (19)$$

onde $\lambda = \tau^{-1}$, e $x!!$ é denotado como o produto de todos números ímpares de 1 a x .

Após o consumo ser estimado, ele é multiplicado pela probabilidade de consumo estimada e, por fim, são calculados os percentis, a média, o desvio padrão, a curtose e a assimetria.

Suposição de consumo diário

Já na suposição de consumo diário é utilizada apenas a segunda parte da metodologia recém descrita.

3.1.4.3. STATISTICAL PROGRAM TO ASSESS DIETARY EXPOSURE (SPADE)

O método SPADE (16–18) realiza a estimação da distribuição do consumo usual de componentes alimentares utilizando a informação da variável idade, assim como seu antecessor, o método AGEMODE (23). Na estimação da distribuição de consumo alimentar usual o método tem características adicionais as apresentadas do Quadro 1:

- O padrão do método é utilizar a idade como covariável na modelagem, entretanto é possível realizar a estimação sem essa informação desde que se faça alguns ajustes,
- Além da idade é possível utilizar outras covariáveis,

- Os resultados podem ser apresentados por unidade de idade (ex. anos ou meses) ou para grupos de idade,
- Podem ser apresentados os intervalos de confiança calculados por *Bootstrap*.

O método tem dois tipos suposição quanto a frequência de consumo na estimação: consumo diário ou consumo episódico. A seguir, foram apresentados os detalhes técnicos da estimação para cada uma das suposições.

O SPADE está implementado no R e utiliza os seguintes pacotes: *mpf* (43), *amer* (44), *lme4* (45), *gamlss* (46).

Suposição de consumo episódico

Para esse tipo de suposição é possível utilizar o modelo que permite identificar nos dados quem são os indivíduos que realmente não consomem o componente alimentar. A identificação desses indivíduos possibilita que o modelo atribua probabilidade de consumo igual a zero para os que não consomem e probabilidade de consumo positiva para os que são consumidores em potencial.

Assim como no NCI e no MSM para esse tipo de suposição o método é separado em duas partes: uma que estima a probabilidade de consumo e outra que estima a quantidade de consumo.

Primeira parte:

Nessa etapa a probabilidade de consumo é estimada em função das covariáveis com um modelo Beta-Binomial.

Segunda parte:

O processo de estimação da quantidade de consumo considerando a suposição de consumo episódico foi detalhado a seguir na Figura 10.

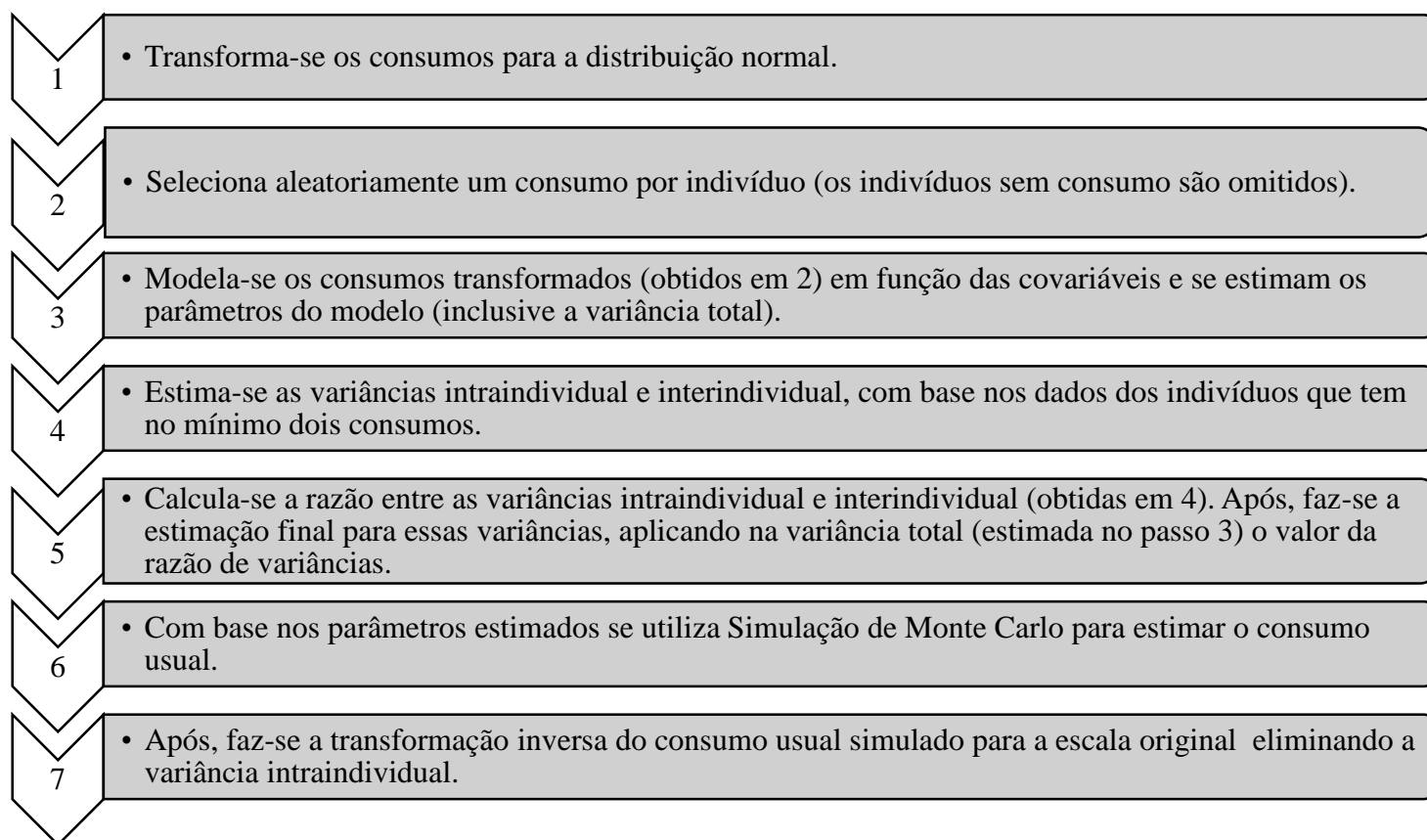


Figura 10: Segunda parte do processo de estimação para a suposição de consumo episódico.

Por fim, a distribuição do consumo alimentar usual é obtida pela multiplicação da probabilidade de consumo pelos dados simulados na escala original.

Suposição de consumo diário

Essa metodologia tem o objetivo de estimar a distribuição de consumo usual de componentes alimentares que são consumidos diariamente. Antes de detalhar as etapas do método, foi apresentada a Figura 11, na qual foi disposto um fluxograma que mostra o processo de estimação.

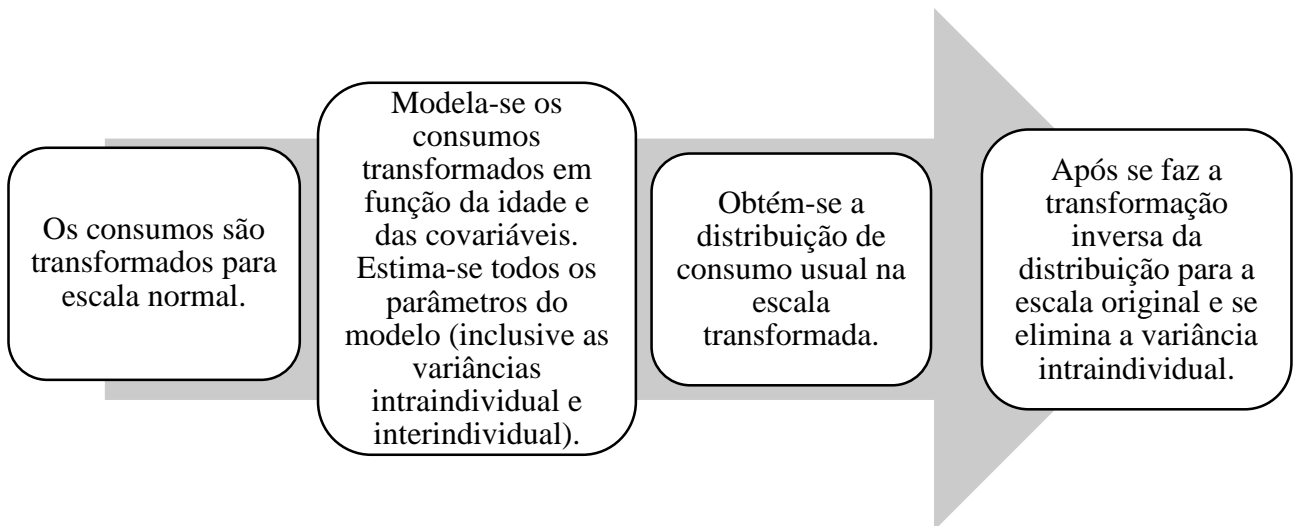


Figura 11: Processo de estimação para a suposição de consumo diário.

Primeiramente os consumos positivos são transformados para a distribuição normal, para isso o modelo utiliza a transformação Box-Cox de um parâmetro (conforme realizado no método NCI). Assim como nos demais métodos essa transformação dos dados não é garantia de que os mesmos tenham distribuição normal. Entretanto, diferentemente dos métodos NCI e MSM, o SPADE possibilita verificar se os dados foram transformados satisfatoriamente para a distribuição normal com base na disponibilização de um gráfico comparativo.

Após a transformação dos dados, o consumo médio individual é modelado com um Modelo Misto (MM) que permite a estimação da variância intraindividual ($\hat{\sigma}_\varepsilon^2$) e da variância interindividual ($\hat{\sigma}_u^2$) conforme a equação a seguir

$$Y_{ij}^* = a + u_i + b(x_i)^p + c(x_i)^q + X_{2i} + \varepsilon_{ij} \quad (20)$$

onde Y_{ij}^* é o consumo transformado do indivíduo i no dia j , a é o intercepto, u_i é o erro aleatório interindividual, x_i é a idade do indivíduo i ; b e c são os coeficientes do modelo para a idade; q e p são as potências do polinômio (que podem assumir valores no conjunto $\{-2; -1; -0,5; 0; 0,5; 1; 2\}$), X_{2i} é o vetor de covariáveis e ε_{ij} é o erro aleatório intraindividual.

Em seguida o método de Quadratura Gaussiana (17) é utilizado para realizar a transformação inversa da distribuição (por unidade de idade) eliminando a variância intraindividual.

3.2. JUSTIFICATIVA DO PRESENTE TRABALHO

Como não há literatura que trate do impacto da escolha das suposições de frequência de consumo, nem do uso de covariáveis sexo e idade na estimação da distribuição do consumo alimentar usual, assim como há poucos estudos que realizaram a avaliação do impacto do QFA como covariável (6 estudos ao total), o presente trabalho avaliou o impacto de diferentes suposições sobre a frequência do consumo (diário ou episódico) e a utilização de covariáveis (QFA, sexo e idade) na estimação da distribuição de consumo alimentar usual.

Essa análise foi realizada para os métodos NCI, MSM e SPADE com base nos dados de Verly et al. (26), que consiste em um conjunto de dados com 20 R24h por indivíduo onde foram mensurados o consumo de diversos componentes alimentares em uma população de residentes na cidade do Rio de Janeiro – Brasil. Para fazer as comparações dos métodos foi utilizada como verdadeira distribuição de consumo a distribuição da média dos 20 R24h individuais dos 302 sujeitos disponíveis.

4. OBJETIVOS

A seguir estão dispostos os objetivos do presente estudo.

4.1. OBJETIVO GERAL

Comparar a acurácia dos métodos de estimação da distribuição do consumo alimentar usual sob diferentes suposições.

4.2. OBJETIVOS ESPECÍFICOS

- Avaliar o impacto da escolha da suposição de frequência alimentar na acurácia dos métodos de estimação da distribuição do consumo alimentar usual;
- Estudar o impacto do uso de covariáveis (QFA, idade e sexo) na acurácia dos métodos de estimação da distribuição do consumo alimentar usual.

REFERÊNCIAS

1. on behalf of the EFCOVAL Consortium, de Boer EJ, Slimani N, van 't Veer P, Boeing H, Feinberg M, et al. The European Food Consumption Validation Project: conclusions and recommendations. *Eur J Clin Nutr.* julho de 2011;65(S1):S102–7.
2. U.S. Department of Agriculture and U.S. Department of Health and Human Services. *Dietary Guidelines for Americans, 2010* [Internet]. 7th Edition, Washington, DC: U.S. Government Printing Office; 2010. Disponível em: <https://health.gov/dietaryguidelines/dga2010/DietaryGuidelines2010.pdf#page=27>
3. Rossum, CTM, Fransen, HP, Verkaik-Kloosterman, J, Buurma-Rethans, EJM, Ocké, MC. Dutch National Food Consumption Survey 2007-2010 [Internet]. National Institute for Public Health and the Environment; 2011 p. 33–7. (Diet of children and adults aged 7 to 69 years). Report No.: 350050006/2011. Disponível em: <https://www.rivm.nl/bibliotheek/rapporten/350050006.pdf>
4. National Research Council (US) Subcommittee on Criteria for Dietary Evaluation. *Nutrient Adequacy: Assessment Using Food Consumption Surveys* [Internet]. Washington (DC): National Academies Press (US); 1986 [citado 18 de agosto de 2018]. Disponível em: <http://www.ncbi.nlm.nih.gov/books/NBK217533/>
5. Dodd KW, Guenther PM, Freedman LS, Subar AF, Kipnis V, Midthune D, et al. Statistical methods for estimating usual intake of nutrients and foods: a review of the theory. *J Am Diet Assoc.* outubro de 2006;106(10):1640–50.
6. Slob W. Modeling long-term exposure of the whole population to chemicals in food. *Risk Anal Off Publ Soc Risk Anal.* outubro de 1993;13(5):525–30.
7. Slob W. Probabilistic dietary exposure assessment taking into account variability in both amount and frequency of consumption. *Food Chem Toxicol Int J Publ Br Ind Biol Res Assoc.* julho de 2006;44(7):933–51.
8. Wallace LA, Duan N, Ziegenfus R. Can Long-Term Exposure Distributions Be Predicted from Short-Term Measurements? *Risk Anal.* fevereiro de 1994;14(1):75–85.
9. Nusser SM, Carriquiry AL, Dodd KW, Fuller WA. A Semiparametric Transformation Approach to Estimating Usual Daily Intake Distributions. *J Am Stat Assoc.* dezembro de 1996;91(436):1440–9.
10. Guenther PM, Kott PS, Carriquiry AL. Development of an approach for estimating usual nutrient intake distributions at the population level. *J Nutr.* junho de 1997;127(6):1106–12.
11. Nusser SM, Fuller WA, Guenther PM. Estimating Usual Dietary Intake Distributions: Adjusting for Measurement Error and Nonnormality in 24-Hour Food Intake Data. In: Lyberg L, Biemer P, Collins M, De Leeuw E, Dippo C, Schwarz N, et al., organizadores. *Survey Measurement and Process Quality* [Internet]. Hoboken, NJ, USA: John Wiley & Sons, Inc.; 2012. p. 689–709. Disponível em: <http://doi.wiley.com/10.1002/9781118490013.ch30>

12. Gay C. Estimation of population distributions of habitual nutrient intake based on a short-run weighed food diary. *British Journal of Nutrition*. 83^o ed 200DC;287–93.
13. Hoffmann K, Boeing H, Dufour A, Volatier JL, Telman J, Virtanen M, et al. Estimating the distribution of usual dietary intake by short-term measurements. *Eur J Clin Nutr*. maio de 2002;56 Suppl 2:S53-62.
14. Haubrock J, Nothlings U, Volatier J-L, Dekkers A, Ocke M, Harttig U, et al. Estimating Usual Food Intake Distributions by Using the Multiple Source Method in the EPIC-Potsdam Calibration Study. *J Nutr*. 23 de março de 2011;141(5):914–20.
15. Harttig U, Haubrock J, Knüppel S, Boeing H. The MSM program: web-based statistics package for estimating usual dietary intake using the Multiple Source Method. *Eur J Clin Nutr*. julho de 2011;65 Suppl 1:S87-91.
16. Souverein OW, Dekkers AL, Geelen A, Haubrock J, de Vries JH, Ocké MC, et al. Comparing four methods to estimate usual intake distributions. *Eur J Clin Nutr*. julho de 2011;65 Suppl 1:S92-101.
17. Dekkers ALM, Slob W. Gaussian Quadrature is an efficient method for the back-transformation in estimating the usual intake distribution when assessing dietary exposure. *Food Chem Toxicol*. outubro de 2012;50(10):3853–61.
18. Dekkers AL, Verkaik-Kloosterman J, van Rossum CT, Ocke MC. SPADE, a New Statistical Program to Estimate Habitual Dietary Intake from Multiple Food Sources and Dietary Supplements. *J Nutr*. 1^o de dezembro de 2014;144(12):2083–91.
19. Tooze JA, Grunwald GK, Jones RH. Analysis of repeated measures data with clumping at zero. *Stat Methods Med Res*. agosto de 2002;11(4):341–55.
20. Tooze JA, Midthune D, Dodd KW, Freedman LS, Krebs-Smith SM, Subar AF, et al. A new statistical method for estimating the usual intake of episodically consumed foods with application to their distribution. *J Am Diet Assoc*. outubro de 2006;106(10):1575–87.
21. Kipnis V, Midthune D, Buckman DW, Dodd KW, Guenther PM, Krebs-Smith SM, et al. Modeling data with excess zeros and measurement error: application to evaluating relationships between episodically consumed foods and health outcomes. *Biometrics*. dezembro de 2009;65(4):1003–10.
22. Tooze JA, Kipnis V, Buckman DW, Carroll RJ, Freedman LS, Guenther PM, et al. A mixed-effects model approach for estimating the distribution of usual intake of nutrients: The NCI method. *Stat Med*. 30 de novembro de 2010;29(27):2857–68.
23. Waijers PMCM, Dekkers ALM, Boer JMA, Boshuizen HC, van Rossum CTM. The potential of AGE MODE, an age-dependent model, to estimate usual intakes and prevalences of inadequate intakes in a population. *J Nutr*. novembro de 2006;136(11):2916–20.
24. Goedhart PW, van der Voet H, Knüppel S, Dekkers ALM, Dodd KW, Boeing H, et al. A comparison by simulation of different methods to estimate the usual intake

- distribution for episodically consumed foods. EFSA - Supporting publications [Internet]. EN299 ed 15 de junho de 2012 [citado 28 de maio de 2018]; Disponível em: <http://doi.wiley.com/10.2903/sp.efsa.2012.EN-299>
25. Laureano GHC, Torman VBL, Crispim SP, Dekkers ALM, Camey SA. Comparison of the ISU, NCI, MSM, and SPADE Methods for Estimating Usual Intake: A Simulation Study of Nutrients Consumed Daily. *Nutrients*. 15 de março de 2016;8(3):166.
 26. Verly-Jr E, Oliveira DCRS, Fisberg RM, Marchioni DML. Performance of statistical methods to correct food intake distribution: comparison between observed and estimated usual intake. *Br J Nutr*. setembro de 2016;116(5):897–903.
 27. Subar AF, Dodd KW, Guenther PM, Kipnis V, Midthune D, McDowell M, et al. The Food Propensity Questionnaire: Concept, Development, and Validation for Use as a Covariate in a Model to Estimate Usual Food Intake. *J Am Diet Assoc*. outubro de 2006;106(10):1556–63.
 28. Verly- E, Fisberg RM, Marchioni DML. Is the food frequency consumption essential as covariate to estimate usual intake of episodically consumed foods? *Eur J Clin Nutr*. novembro de 2012;66(11):1254–8.
 29. Ost C, De Ridder KAA, Tafforeau J, Van Oyen H. The added value of food frequency questionnaire (FFQ) information to estimate the usual food intake based on repeated 24-hour recalls. *Arch Public Health* [Internet]. dezembro de 2017 [citado 22 de julho de 2018];75(1). Disponível em: <http://archpublichealth.biomedcentral.com/articles/10.1186/s13690-017-0214-8>
 30. Willett W. *Nutritional epidemiology*. Third edition. Oxford ; New York: Oxford University Press; 2013. 529 p. (Monographs in epidemiology and biostatistics).
 31. Applied Research: Cancer Control and Population Sciences.Measurement Error [Internet]. Webinar Series. [citado 20 de agosto de 2018]. Disponível em: <http://appliedresearch.cancer.gov/measurementerror/>
 32. Guenther PM, Kirkpatrick SI, Reedy J, Krebs-Smith SM, Buckman DW, Dodd KW, et al. The Healthy Eating Index-2010 is a valid and reliable measure of diet quality according to the 2010 Dietary Guidelines for Americans. *J Nutr*. março de 2014;144(3):399–407.
 33. Usual Dietary Intakes: Food Intakes, U.S. Population, 2007-10 [Internet]. Disponível em: <https://epi.grants.cancer.gov/diet/usualintakes/pop/2007-10/>
 34. Crispim SP, de Vries JHM, Geelen A, Souverein OW, Hulshof PJM, Lafay L, et al. Two non-consecutive 24 h recalls using EPIC-Soft software are sufficiently valid for comparing protein and potassium intake between five European centres – results from the European Food Consumption Validation (EFCOVAL) study. *Br J Nutr*. fevereiro de 2011;105(03):447–58.
 35. Yang Q, Zhang Z, Gregg EW, Flanders WD, Merritt R, Hu FB. Added Sugar Intake and Cardiovascular Diseases Mortality Among US Adults. *JAMA Intern Med*. 1º de abril de 2014;174(4):516.

36. Herrick, KA R LM, Parsons, R D KW. Estimating usual dietary intake from National Health and Nutrition Examination Survey data using the National Cancer Institute method. National Center for Health Statistics. 2018. Report No.: Vital Health Stat 2(178).
37. Subar AF, Thompson FE, Kipnis V, Midthune D, Hurwitz P, McNutt S, et al. Comparative validation of the Block, Willett, and National Cancer Institute food frequency questionnaires : the Eating at America's Table Study. *Am J Epidemiol.* 15 de dezembro de 2001;154(12):1089–99.
38. Buck RJ, Hammerstrom KA, Ryan PB. Estimating long-term exposures from short-term measurements. *J Expo Anal Environ Epidemiol.* setembro de 1995;5(3):359–73.
39. SAS Institute. Base SAS 9.3 procedures guide [Internet]. Cary, N.C.: SAS Institute; 2011 [citado 29 de maio de 2018]. Disponível em: <http://www.books24x7.com/marc.asp?bookid=44483>
40. Pinheiro JC, Bates DM. Approximations to the Log-Likelihood Function in the Nonlinear Mixed-Effects Model. *J Comput Graph Stat.* 1º de março de 1995;4(1):12–35.
41. Box GEP, Cox DR. An Analysis of Transformations. Series B (Methodological). 1964;211–52.
42. Usual Dietary Intakes: Details of the NCI Method [Internet]. Disponível em: <https://epi.grants.cancer.gov/diet/usualintakes/details.html>
43. Ambler G, Benner A. MFP: Multivariable Fractional Polynomials. 2010; Disponível em: <http://cran.r-project.org/web/packages/mfp/>
44. Scheipl F. AMER: Additive mixed models with lme4. 2011; Disponível em: <http://cran.r-project.org/web/packages/amer/amer.pdf>
45. Bates DM, Maechler M, Bolker B. LME4: Linear mixed-effects models using S4 classes. setembro de 2013; Disponível em: <http://cran.r-project.org/web/packages/lme4/>
46. Stasinopoulos DM, Rigby RA. Generalized Additive Models for Location Scale and Shape (GAMLSS) in R. *Journal of Statistical Software.* 23º ed dezembro de 2007;

7. CONCLUSÕES E CONSIDERAÇÕES FINAIS

Esse trabalho teve como objetivo preencher lacunas existentes na literatura de nutrição epidemiológica no que se refere ao impacto da escolha da suposição de frequência de consumo alimentar (diário ou episódico), assim como da utilização de covariáveis (QFA, sexo e idade) na estimação da distribuição de consumo alimentar usual, chegando-se às seguintes recomendações:

- Utilizar a suposição de frequência de consumo diário na estimação da distribuição de consumo alimentar de componentes alimentares de frequência de consumo diária, independentemente do método;
- Utilizar a suposição de frequência de consumo episódico em componentes de frequência de consumo episódico de nível intermediário:
 - Emprego do método MSM na estimação dos percentis 5 e 10;
 - Uso do NCI, MSM ou SPADE na estimação dos demais parâmetros da distribuição.
- Utilizar da suposição de frequência de consumo episódico em componentes de frequência de consumo episódico de nível baixo:
 - Emprego do método MSM na estimação dos percentis menores ou iguais a 25;
 - Uso dos métodos SPADE e NCI para estimativa da média e dos percentis acima de 25.

Conclui-se que, em relação ao uso do QFA, apesar de ter sido apurado discreto aumento de acurácia em algumas situações, principalmente nos métodos MSM e SPADE, o custo-benefício da coleta dessas informações deve ser avaliado, visto que a elaboração e aplicação do QFA demanda investimentos que podem não compensar as melhoras

obtidas.

Já quando se trata da utilização das covariáveis sexo e idade, recomenda-se o uso de todos os métodos sem covariáveis na estimação de componentes alimentares de frequência de consumo diário e episódico de nível intermediário. Todavia, nos componentes alimentares com frequência de consumo episódico de nível baixo, recomenda-se a utilização do método MSM com ajuste por sexo e idade e dos métodos NCI e SPADE sem esse ajuste.

Os achados desse estudo não corroboraram em totalidade os da literatura atual no que se refere a escolha das suposições, uma vez que nos dados analisados não houveram componentes alimentares com percentual de zeros nos R24h entre 5% e 10%, entretanto, foi possível verificar que para componentes alimentares com até 0,1% de R24h iguais a zero, a suposição com maior acurácia é a de consumo diário e, quando no mínimo 20% de R24h são iguais a zero, a utilização da suposição de consumo episódico é mais adequada.

Já para o uso do QFA como informação adicional na modelagem os resultados confirmam os achados atuais da literatura no que se refere ao melhor desempenho do método SPADE - quando utilizado a informação de quem nunca consome o componente alimentar - na estimação de percentis inferiores, como também, da falta de impacto da utilização do QFA como covariável no NCI. Além disso, o presente trabalho trouxe novos resultados a respeito do impacto do uso do QFA como covariável no método MSM.

Embora tenha sido possível realizar uma série de recomendações, ainda são necessárias mais investigações a respeito da suposição de frequência de consumo alimentar e do uso de covariáveis na estimação dos métodos, para que se possa ter maior confiança na escolha dos métodos, assim como saber se é imperativo que se utilize covariáveis na estimação da distribuição de consumo alimentar usual. Em vista disso,

estudos futuros poderiam investigar a influência da suposição de frequência de consumo em componentes alimentares com consumos iguais a zero entre 0,1% e 20% nos R24h, como também a análise do impacto da utilização de outras covariáveis na estimação dos métodos, além de repetir as análises realizadas em outros conjuntos de dados.