

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



DISSERTAÇÃO DE MESTRADO

Comparação de métodos para estimação da distribuição do consumo alimentar usual.

Greice Helen da Costa Laureano

Orientadora: Prof. Dra. Suzi Alves Camey

Porto Alegre, Fevereiro de 2014.

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



DISSERTAÇÃO DE MESTRADO

Comparação de métodos para estimação da distribuição do consumo alimentar usual.

Greice Helen da Costa Laureano

Orientador: Prof. Dra. Suzi Alves Camey

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

Porto Alegre, Brasil.
2014

CIP - Catalogação na Publicação

Laureano, Greice Helen da Costa
Comparação de métodos para estimação da distribuição
do consumo alimentar usual. / Greice Helen da Costa
Laureano. -- 2014.
64 f.

Orientador: Suzi Alves Camey.

Dissertação (Mestrado) -- Universidade Federal do
Rio Grande do Sul, Faculdade de Medicina, Programa
de Pós-Graduação em Epidemiologia, Porto Alegre, BR-
RS, 2014.

1. Estimação. 2. Distribuição de consumo alimentar.
3. MEID. I. Camey, Suzi Alves, orient. II. Título.

BANCA EXAMINADORA

Prof. Dr. Alvaro Vigo, Programa de Pós-graduação em Epidemiologia, UFRGS.
Prof. Dra. Vivian Cristine Luft, Programa de Pós-graduação em Epidemiologia, UFRGS.
Prof. Dra. Sandra Patricia Crispim, Faculdade de Saúde Pública - Departamento de Nutrição, USP.

DEDICATÓRIA

Dedico este trabalho ao Rodrigo Coster que me apoiou e deu suporte em todas as etapas.

AGRADECIMENTOS

Agradeço em primeiro lugar a minha família que sempre me incentivou a estudar. Também agradeço a família Coster pelo apoio e carinho, às professoras Suzi e Vanessa que me guiaram na realização deste trabalho, aos meus amigos e, em especial, ao meu namorado Rodrigo Coster.

SUMÁRIO

Abreviaturas e siglas.....	7
Resumo	8
Abstract.....	9
1. APRESENTAÇÃO.....	10
2. INTRODUÇÃO.....	11
3. REVISÃO DE LITERATURA	14
3.1. MÉTODOS DE ESTIMAÇÃO DO CONSUMO ALIMENTAR USUAL... 14	
3.1.1. NATIONAL CANCER INSTITUTE (NCI)	16
3.1.2. MULTIPLE SOURCE METHOD (MSM)	21
3.1.3. IOWA STATE UNIVERSITY METHOD (ISU).....	23
3.1.4. STATISTICAL PROGRAM TO ASSESS DIETARY EXPOSURE (SPADE).....	27
3.1.5. GAY	30
3.1.6. SIMPLIFIED NUSSER (S-NUSSER)	31
3.1.7. STEM.I E STEM.II	34
3.1.8. WALLACE.....	36
3.1.9. AGEMODE	37
3.2. COMPARAÇÃO DE MÉTODOS.....	40
4. OBJETIVO	44
5. ARTIGO	45
6. CONCLUSÕES E CONSIDERAÇÕES FINAIS	60
7. REFERÊNCIAS BIBLIOGRÁFICAS	61

Abreviaturas e siglas

ANOVA – *Analysis of Variance*

GEE – *Generalized Estimation Equation*

ISU – *Iowa State University Method*

NCI – *National Cancer Institute*

MEID – *Method for Estimate Intake Distribution*

MMNL – *Modelo Misto Não Linear*

MSM – *Multiple Source Method*

QFA – *Questionário de Frequência Alimentar*

R24h – *Recordatório Alimentar de 24 horas*

SPADE – *Statistical Program to Assess Dietary Exposure*

Resumo

Já há inúmeros métodos de estimação da distribuição do consumo alimentar usual. De uma forma geral, eles são de difícil entendimento e implementados em *softwares* pagos. Em vista disso, este trabalho propõe um novo método de estimação da distribuição do consumo alimentar usual que foi intitulado como MEID (*Method for Estimate Intake Distribution*).

Neste trabalho esse método foi comparado com os métodos SPADE (*Statistical Program to Assess Dietary Exposure*), NCI (*National Cancer Institute*) e MSM (*Multiple Source Method*), por meio de um estudo de simulação.

Para cada método foram estimados a média e os percentis da distribuição do consumo usual e foram utilizadas como medidas de acurácia o vício e o vício relativo, já para medir a precisão foi utilizado o Erro Quadrático Médio (EQM). Salienta-se que o método proposto neste trabalho utilizou o *software* livre R em todo seu escopo.

Os resultados mostraram que os métodos NCI, SPADE e MSM realizaram estimações parecidas e obtiveram EQM e vícios superiores ao MEID, principalmente quando se tratava do percentil 95.

Em suma, verificou-se que o estimador MEID foi mais acurado e preciso na estimação da distribuição do consumo alimentar usual quando comparado aos demais, como também se mostrou mais simples e de interface implementável em qualquer *software* com linguagem similar ao do *software* R.

Abstract

There are already numerous methods for estimating the distribution of usual dietary intake. In general, they are difficult to understand and implemented in paid software. In this light, this paper proposes a new method for estimating the distribution of usual dietary intake: MEID (Method for Intake Distribution Estimate).

In this work this method was compared with the SPADE (Statistical Program to Assess Dietary Exposure), NCI (National Cancer Institute) and MSM (Multiple Source Method) methods, by means of a simulation study.

For each method, the mean and the percentiles of usual intake were estimated, and for accuracy measurements, the bias and relative bias were used. The precision was measured using the Mean Squared Error (MSE). The proposed method in this paper uses the free software R.

The results have shown that the NCI, MSM and SPADE methods performed similar estimations, and obtained MSE and bias higher than MEID, especially in the 95th percentile.

In short, it was found that the MEID estimator was more accurate and precise in estimating the distribution of usual dietary intake, when compared to other estimators, while also being simpler than those. It was also shown that it's implementable in any language similar to the R software.

1. APRESENTAÇÃO

Este trabalho consiste na dissertação de mestrado intitulada “Comparação de métodos para estimação da distribuição do consumo alimentar usual”, apresentada ao Programa de Pós-Graduação em Epidemiologia da Universidade Federal do Rio Grande do Sul, em 13 de Fevereiro de 2014. O trabalho é apresentado em três partes, na ordem que segue:

1. Introdução, Revisão da Literatura e Objetivos
2. Artigo
3. Conclusões e Considerações Finais.

Documentos de apoio estão apresentados nos anexos.

2. INTRODUÇÃO

A mensuração do consumo alimentar usual é assunto vigente na área da saúde, uma vez que várias doenças são influenciadas, ou até mesmo causadas, pelos hábitos alimentares da população (1).

O Questionário de Frequência Alimentar (QFA) e o Recordatório Alimentar de 24 horas (R24h) são utilizados como um dos meios de obter estimativas de consumo de alimentos e nutrientes na população. No QFA, os sujeitos do estudo respondem a respeito da frequência em que consomem determinados alimentos em um período de tempo, por exemplo, a quantidade de carne vermelha consumida semanalmente, já no R24h, é relatado tudo que foi consumido durante o dia. Existem também outros métodos de mensuração do consumo alimentar como os registros alimentares e pesagem de alimentos que não foram detalhados neste trabalho.

No Brasil, a utilização dos métodos de mensuração alimentar, em geral, é realizada sem corrigir os vieses de mensuração que lhes são inerente, em contrapartida estudos realizados em outros países indicam a necessidade da correção dos mesmos (2–10). Um exemplo de consequência do viés de mensuração é encontrado na estimação da distribuição do consumo alimentar, onde pode ocorrer sub ou superestimação do percentual de pessoas que apresentam um consumo inadequado de um alimento ou nutriente.

O controle dos vieses tem duas abordagens e serve para o controle dos vieses sistemáticos ou dos vieses aleatórios. A primeira abordagem é realizada no momento da aplicação do QFA e do R24h (essa abordagem é referente ao controle de vieses sistemáticos oriundos dos métodos de mensuração de consumo), já na segunda abordagem é realizada a correção dos vieses após a aplicação do QFA ou do R24h, uma vez que os métodos de correção são realizados através de modelos estatísticos.

Dentro dos métodos de correção posterior a aplicação existe aqueles que são aplicados na estimação da distribuição do consumo alimentar usual, que tem como objetivo estimar o consumo alimentar de populações realizando a correção do viés aleatório inerente aos métodos de mensuração de consumo (11). Essa correção é feita no ajuste do modelo estatístico nos dados de consumo alimentar, uma vez que nesse ajuste se retira a variabilidade aleatória atribuída ao indivíduo, para que se possa obter a distribuição de consumo usual da população.

Alguns desses métodos estão implementados nos seguintes aplicativos computacionais: STEM.I (12) e STEM.II (13)/Genérico; Wallace (14)/Programa *Small*; ISU (15–17)/ Versão SIDE no SAS e Versão C-SIDE no UNIX; Gay (18)/ Genérico; S-Nusser (19)/SAS; MSM (20)/Web; SPADE (21–23)/R; NCI (24–27)/SAS e AGE MODE (28)/ S-PLUS.

Os métodos citados acima, de uma forma geral, são de difícil entendimento para o público devido à complexa metodologia que os envolve e também são limitados ao uso de *softwares* específicos e/ou pagos (o detalhamento deles foi apresentado na sequência).

Em virtude disso, este trabalho propõe um novo método de estimação da distribuição do consumo alimentar usual: MEID (*Method for Estimate Intake Distribution*). A proposição desse novo método tem como objetivos: usar um modelo metodologicamente mais simples e, portanto, de fácil entendimento e também fazer o uso do *software* livre R.

Alguns autores realizaram estudos em que comparavam métodos de estimação da distribuição do consumo alimentar usual (11,15,16,19,22,23,27). Destacando-se Souverein *et al.* (23) que realizaram um estudo de simulação comparando os métodos SPADE (*Statistical Program to Assess Dietary Exposure*), NCI (*National Cancer*

Institute), MSM (*Multiple Source Method*) e ISU (*Iowa State University Method*), com o objetivo de averiguar o quanto as distribuições estimadas de consumo alimentar usual desviam da verdadeira distribuição.

Contudo, houve limitações neste estudo, como, por exemplo, a utilização de 100 replicações para cada cenário, uma vez que geralmente estudos de simulação trabalham com 1.000 ou 10.000 replicações para poder dar estabilidade às estimativas. Como consequência, não foi possível concluir sobre qual o melhor método em alguns cenários e percentis devido à grande variabilidade existente nas estimativas.

Neste trabalho se desejou avaliar as características da estimação da distribuição alimentar usual de um alimento ou nutriente. Por isso foi realizado um estudo de simulação baseado na abordagem de Souverein *et al.* (23), onde foram comparados os métodos SPADE, NCI e MSM ao novo método proposto. Além da diferença nos métodos comparados, foi utilizado um número maior de replicações, dando enfoque aos cenários que apresentaram maior incerteza nas conclusões. Como referência de comparação foi utilizada a distribuição Box-Cox (29).

3. REVISÃO DE LITERATURA

Nas sessões a seguir foram detalhados os métodos de estimação da distribuição do consumo alimentar usual e também foram comentados alguns trabalhos que compararam métodos de estimação.

3.1. MÉTODOS DE ESTIMAÇÃO DO CONSUMO ALIMENTAR USUAL

Nessa sessão foram apresentados alguns métodos para estimação da distribuição do consumo alimentar usual encontrados na literatura. Tais métodos foram propostos com a finalidade de fazer a estimação para populações em relação ao consumo de alimentos e nutrientes se valendo dos R24h. Alguns desses métodos são citados no Quadro 1 a seguir.

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

Método	Ano de publicação	Software	Tipo	Permite uso de covariáveis?	Modela o consumo episódico?	Permite a modelagem de mais de um item alimentar?	Planos Complexos
NCI (21–24)	2009	SAS	Pago	Sim	Sim	Sim	Sim
MSM (19,32)	2011	Web	Livre	Sim	Sim	Sim	Sim
ISU (14–16)	1996	Versão SIDE no SAS, Versão C-SIDE e PC-SIDE.	Pago	Sim	Sim	Não	Sim
SPADE (21–23)	2010	R	Livre	Sim	Sim	Não	Não
Gay (17)	1994	Programa que tenha a ANOVA* permita simulação.	-	Não	Não	Não	Não
S-Nusser (18)	2002	SAS	Pago	Sim	Não	Não	Sim
STEM.I e II (10,11)	1996	Programa que tenha a ANOVA*	-	Sim	Sim	Não	Não

Método	Ano de publicação	<i>Software</i>	Tipo	Permite uso de covariáveis?	Modela o consumo episódico?	Permite a modelagem de mais de um item alimentar?	Planos Complexos
Wallace (12)	1999	Programa <i>Small</i>	-	-	Não	Não	Não
AGE MODE (25)	2006	S-PLUS	Pago	Sim	Não	Não	Não

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

Os dados de consumo alimentar, em geral, têm distribuições assimétricas, o que impossibilita a utilização de diversos métodos para a estimação do consumo. Uma tática utilizada em larga escala é a transformação dos dados para uma distribuição normal, para que assim se possam utilizar métodos que exigem o desfecho e/ou os resíduos do modelo normalmente distribuídos.

Em geral, os métodos têm o objetivo de estimar a média e a variância de consumo da população para poder estimar a distribuição do consumo usual. Assim, após a obtenção das estimativas na escala normal, é feita a transformação inversa dos dados para a escala original e então, se calcula a média e os percentis das mesmas.

Os métodos de estimação se diferenciam em relação à forma de estimação da média e da variância e a forma de transformação dos dados, mas em geral, eles têm as seguintes pressuposições (12–21,24–28,30):

- 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 abordadas as metodologias que se referem ao consumo usual de alimentos, ou seja, as que permitem a estimação da distribuição do consumo esporádico de alimentos não foram detalhadas.

Para facilitar a leitura foi referido a “estimação de alimentos ou nutrientes” como “estimação dos itens alimentares”.

3.1.1. NATIONAL CANCER INSTITUTE (NCI)

O NCI foi proposto por Tooze *et al.* (24,25,27) e Kipnis *et al.* (26) para estimar a distribuição do consumo alimentar usual e esporádico de itens alimentares a partir da aplicação de no mínimo dois R24h's. O método foi adaptado ao passar dos anos e neste trabalho foi abordada a versão mais atual de 2013 para estimação da distribuição de consumo de um item alimentar.

Esse método é composto por duas macros do SAS: a primeira é intitulada MIXTRAN, que ajusta um Modelo Misto Não Linear (MMNL) aos dados de consumo alimentar, com o intuito de estimar os parâmetros necessários para a segunda macro, intitulada como DISTRIB, que utiliza o método de simulação Monte Carlo para estimar a distribuição de consumo usual da população (26,27).

O NCI permite que algumas covariáveis sejam adicionadas no modelo, como por exemplo, a idade, o IMC e informações adicionais de um questionário de propensão alimentar (QFP). As macros do SAS podem ser encontradas no site: <http://riskfactor.cancer.gov/diet/usualintakes/>.

A seguir detalhes foram fornecidos sobre as duas macros.

A macro MIXTRAN estima os parâmetros referentes à quantidade consumida do item (o valor esperado de consumo - considerando as covariáveis - e variâncias intraindividual e interindividual).

Nessa etapa se utiliza, inicialmente, o método Equações de Estimação Generalizadas (GEE, do inglês *Generalized Estimating Equation*) para estimar os parâmetros (o valor esperado de consumo e as variâncias intraindividual e interindividual).

Posteriormente, um MMNL é ajustado - utilizando como valores iniciais os valores dos parâmetros estimados no GEE - para reestimar os valores das variâncias intraindividual e interindividual. O ajuste do MMNL é realizado via otimização *Quasi-Newton*, que utiliza a verossimilhança aproximada pelo método de *Quadratura Gaussiana Adaptativa* (31).

Para facilitar o entendimento, antes de se detalhar o método, foi apresentada a Figura 1, na qual foi disposto um organograma que mostra de forma simplificada o processo de estimação.

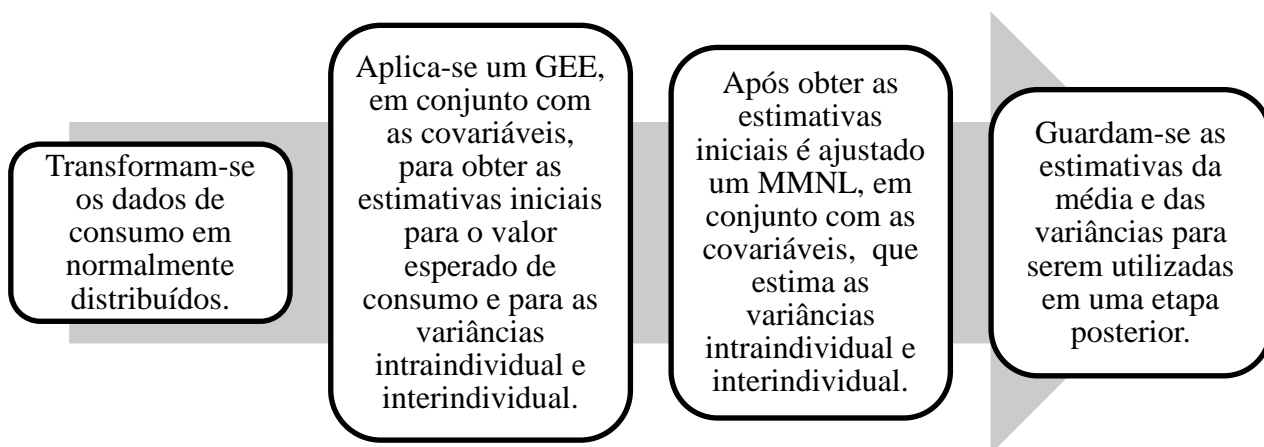


Figura 1. Processo de estimação da média de consumo geral e das variâncias para o consumo do item alimentar do método NCI.

Como mostrado na Figura 1 os dados de consumo são transformados para a escala normal pelo método. Considerando a variável aleatória Y_{ij} que representa a quantidade consumida do item alimentar, pelo sujeito i no tempo j , tem-se que a transformação dos dados é realizada por meio da transformação Box-Cox de um parâmetro (29) que é definida como

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

onde Y_{ij}^* são os dados de consumo transformados para a distribuição normal 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).

Pode-se então explicar o consumo transformado através de uma média individual (μ_i^*) e um desvio aleatório (ε_{ij}) que captará as variações diárias de alimentação dos indivíduos. Então se tem que

$$Y_{ij}^* = \mu_i^* + \varepsilon_{ij}, \quad \varepsilon_{ij} \sim N(0, \sigma_\varepsilon^2) \quad (2)$$

onde se assume que os erros intraindividuais (ε_{ij}) são aditivos, não correlacionados com média individual (μ_i^*) e independentes uns dos outros. Tem-se também que σ_ε^2 é a variância do efeito intraindividual.

Quando se opta pelo uso de covariáveis, tem-se que o vetor de covariáveis (X_i') pode ser incorporado na estimação da média individual conforme mostra a equação

$$\mu_i^* = X_i' \beta + u_i, \quad u_i \sim N(0, \sigma_u^2) \quad (3)$$

onde u_i é o erro da quantidade consumida que capta características individuais de consumo não explicadas pelas covariáveis (ou erros interindividuais) e σ_u^2 é a sua variância. Reunindo as equações (2) e (3) se tem que

$$Y_{ij}^* = X_i' \beta + u_i + \varepsilon_{ij} \quad (4)$$

Combinando as equações (1) e (4) se tem que o MMNL para o R24h é

$$g(Y_{ij}, \lambda) = X_i' \beta + u_i + \varepsilon_{ij} \quad (5)$$

onde $g(Y_{ij}, \lambda) = Y_{ij}^*$, ou seja, a função que transforma os consumos originais para escala normal.

Como dito anteriormente o MMNL faz a estimação das variâncias σ_u^2 e σ_ε^2 via otimização *Quasi-Newton* que utiliza a verossimilhança aproximada pelo método de Quadratura *Gaussiana Adaptativa* (31). Com isso se tem que a contribuição do sujeito i

para essa verossimilhança é dada por

$$L_i(\lambda, \beta, \sigma_u, \sigma_\varepsilon; y_{i1}, \dots, y_{im}) = \int_{u_i} \prod_{j=1}^m g(y_{ij} | \lambda, \beta, u_i, \sigma_\varepsilon) f(u_i | \sigma_u) du_i,$$

$$g(y_{ij} | \lambda, \beta, u_i, \sigma_\varepsilon) = \frac{1}{\sqrt{2\pi\sigma_\varepsilon^2}} e^{-\left[\frac{\left(\frac{y_{ij}^\lambda - 1}{\lambda - X_i' \beta - u_i}\right)^2}{2\sigma_\varepsilon^2}\right]} y_{ij}^{\lambda-1} \quad (6)$$

onde m é o número de recordatórios, g é a função de densidade do consumo observado y_{ij} , para cada sujeito i em cada dia j . Enquanto f é a função densidade de u_i .

Após a otimização da verossimilhança se estima σ_u^2 e σ_ε^2 . Por fim, a macro guarda os valores de $\hat{\beta}$ - obtido no GEE - e das variâncias $\hat{\sigma}_u^2$ e $\hat{\sigma}_\varepsilon^2$ - obtidas no MMNL - para posteriormente serem utilizadas pela macro DISTRIB.

Essa parte do método tem como objetivo final estimar a média e os percentis da distribuição do consumo usual. Em virtude disso a macro DISTRIB utiliza simulação de Monte Carlo para simular o consumo de k recordatórios, fazendo o uso das estimativas obtidas pela macro MIXTRAN.

Para facilitar o entendimento, antes de se detalhar essa etapa do método, foi apresentada a Figura 2, na qual foi disposto um organograma que mostra de forma simplificada o processo de estimação.

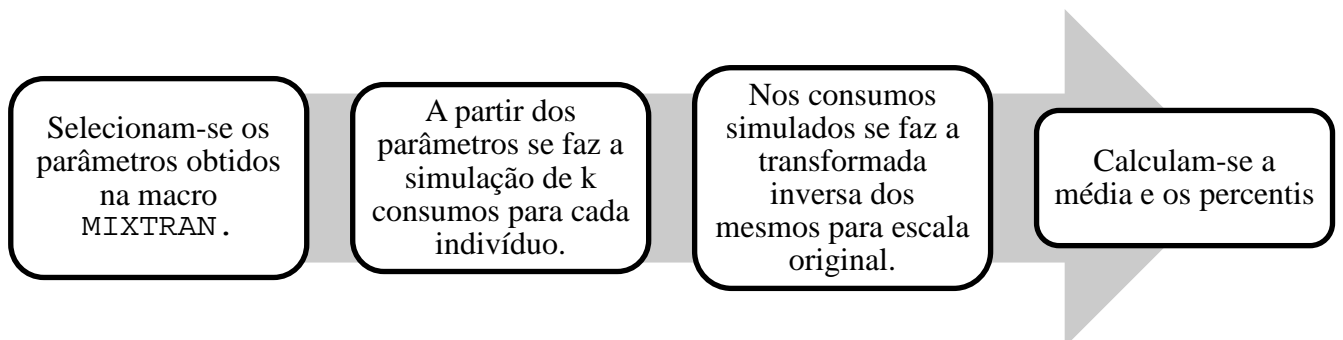


Figura 2. Processo de estimação da distribuição do consumo alimentar usual do método NCI.

O método, em primeiro lugar, calcula $X_i' \hat{\beta}$ para cada sujeito amostrado i (estes

valores representam o valor médio do consumo do item alimentar para cada indivíduo apenas em função dos seus valores de covariáveis). Após se simulam k realizações de u_l , que possui distribuição $N(0, \sigma_u^2)$, e se aplica a fórmula $\mu_l^* = X_l' \hat{\beta} + u_l$ onde $l = 1, \dots, kn$; $X_l = X_i$ para $l = (i - 1)k + 1, \dots, ik$; k é o número de valores simulados e n é o número de indivíduos no estudo.

Com isso, tem-se que μ_l^* são k pseudos consumos para cada indivíduo na escala transformada, obtendo-se um banco de dados simulados com n linhas e k colunas. Segundo os autores, esse conjunto de dados é uma amostra representativa para o consumo usual transformado da população.

Após a criação do banco de dados simulados é utilizada a transformação inversa para trazer os dados para escala original. Essa transformação faz uso da aproximação via série de Taylor, conforme indica o artigo de Tooze *et al.* (27), da seguinte forma

$$T_l \cong g^{-1}(\mu_l^*, \lambda) + \frac{1}{2} \sigma_\varepsilon^2 \frac{\partial^2 \{g^{-1}(\mu_l^*, \lambda)\}}{\partial \mu_l^{*2}} = (\mu_l^* \lambda + 1)^{1/\lambda} + \frac{1}{2} \sigma_\varepsilon^2 (1 - \lambda) (\mu_l^* \lambda + 1)^{1/\lambda - 2} \quad (7)$$

onde T_l é a aproximação de Taylor, g^{-1} é a função inversa de μ_l^* e λ .

Por fim, no conjunto de valores transformados para escala original, calculam-se empiricamente por interpolação linear os percentis e a média amostral.

Os autores argumentam que não utilizam a variância intraindividual na estimação dos percentis, pois eles assumem a definição de que ela não contribui no consumo a longo prazo (25).

Os autores salientam que, quando a variabilidade intraindividual é muito maior que a variabilidade interindividual, a transformação inversa via série de Taylor pode não funcionar bem. Por isso, nesses casos eles sugerem a utilização do método de transformação inversa utilizado pelo método ISU (15).

3.1.2. MULTIPLE SOURCE METHOD (MSM)

O Método MSM foi proposto por Haubrock *et al.* (20) e Harttig *et al.* (32) e está disponível no site <https://msm.dife.de/>. Assim como no NCI, as informações de sexo, idade e QFA podem ser incluídas como covariáveis em duas partes da modelagem.

Para fazer as estimações os autores fazem duas suposições:

- O R24h fornece mensurações não viciadas para o consumo usual.
- Os dados transformados podem ser decompostos em dois componentes independentes (variação intraindividual e interindividual).

Antes de detalhar as etapas do método, foi apresentada a Figura 3, na qual está disposto um organograma que mostra de forma simplificada do processo de estimação.

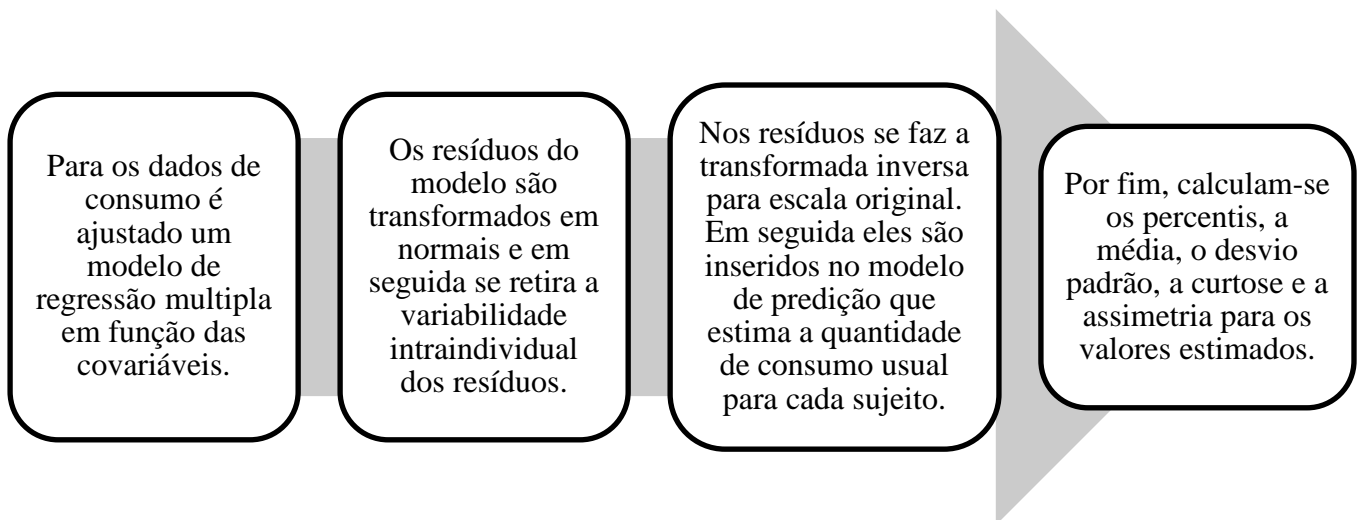


Figura 3. Processo de estimação da distribuição do consumo alimentar usual do método MSM.

Seja a variável aleatória Y_{ij} que representa a quantidade consumida do sujeito i no tempo j , tem-se que o método MSM ajusta uma regressão linear múltipla para os dados de consumo considerando o vetor de características individuais ϱ (frequência de sexo, idade, sexo e idade, e QFA). Fazendo com que a equação da regressão seja definida por

$$Y_{ij} = M_{i/\rho} + R_{ij} \quad (8)$$

onde $M_{i/\rho}$ é a previsão do modelo para o consumo diário do indivíduo i e R_{ij} é o correspondente residual.

Logo em seguida, os resíduos do modelo são transformados para uma distribuição normal pela transformada de Box-Cox de dois parâmetros (29):

$$f(R_{ij}) = \begin{cases} ((R_{ij} + \omega)^\tau - 1)\tau^{-1} & \text{se } \tau \neq 0 \\ \ln(R_{ij} + \omega) & \text{se } \tau = 0 \end{cases} \quad (9)$$

onde $f(R_{ij})$ é a função que transforma os resíduos em normalmente distribuídos, $\omega > R_{ij_{min}}$ e τ é o valor recíproco da integral positiva.

Os valores de ω e τ são determinados por maximização de Shapiro-Wilk, que é realizada em uma macro do SAS. Nessa macro, τ é definido no intervalo $[1, \frac{1}{2}, \frac{1}{3}, \frac{1}{4}, \dots, \frac{1}{10}, 0]$ e ω varia ao longo do mesmo intervalo multiplicado pela média dos dados originais.

Após a transformação dos resíduos é utilizado o modelo de mensuração de erros

$$X_{ij} = f(R_{ij}) = T_i + \varepsilon_{ij} \quad (10)$$

onde T_i é a variável aleatória com esperança E_{T_i} (denotada como valor esperado para o indivíduo i) e ε_{ij} é o termo do erro que inclui a mensuração aleatória da variação diária do consumo. Esses dois termos (T_i e ε_{ij}) são supostos independentes entre si e têm variâncias σ_u^2 e σ_ε^2 , respectivamente. Os autores assumem que ε_{ij} é normalmente distribuído com média zero e variância σ_ε^2 .

Para estimar σ_u^2 , que reflete a variação interindividual na escala transformada dos resíduos, os autores estimam σ_ε^2 e a variância $\sigma_{\bar{x}}^2$ da média individual \bar{X}_i e, em seguida, aplicam a fórmula

$$\hat{\sigma}_u^2 = \hat{\sigma}_{\bar{x}}^2 - \frac{\hat{\sigma}_\varepsilon^2}{k} \quad (11)$$

onde k é o número de dias em que o indivíduo consumiu ($k > 0$).

Por conseguinte, passa-se de uma média individual \bar{X}_i , para uma média geral $\bar{X}_.$ pela redução da variância observada em uma variância estimada $\hat{\sigma}_T^2$.

Posteriormente os novos resíduos são obtidos pela equação

$$\hat{T}_i = \frac{\hat{\sigma}_u}{\hat{\sigma}_X} [\bar{X}_i - \bar{X}_.] + \bar{X}_. \quad (12)$$

Por fim eles são adicionados ao modelo de predição, que realiza a transformada inversa dos resultados para escala original e se obtém a estimativa de consumo usual para o indivíduo i - conforme mostra a equação

$$Y_i^* = M_{i/\rho} + f_{retrans}(\hat{T}_i) \quad (13)$$

onde $f_{retrans}(\hat{T}_i)$ é a transformação inversa dos resíduos.

A transformação inversa é obtida pela inversa da função f (vide equação 9) integrada na distribuição do erro, com ω e τ estimados conforme mostra a equação

$$f_{retrans}(t) = \lambda^{-\lambda} \sum_{s=0}^{int(\lambda/2)} \binom{\lambda}{2s} (t + \lambda)^{\lambda-2s} \sigma_\epsilon^{2s} (2s - 1)!! - \omega \quad (14)$$

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

A equação (14) pode ser matematicamente derivada por integração numérica (19). Tem-se também que para planejamentos desequilibrados pode ser incluso um fator de correção na transformação inversa.

Após o consumo usual ser estimado, são calculados os percentis, a média, o desvio padrão, a curtose e a assimetria desses valores estimados.

3.1.3. IOWA STATE UNIVERSITY METHOD (ISU)

O método ISU têm duas versões: uma definida para estimação do consumo usual baseada em modelo complexo em dois estágios – que foi intitulado neste trabalho de método de ISU^u e a outra que se baseia no método intitulado pelos autores de *Best*

Power que foi criado com a intenção de ser uma versão mais simples que o ISU^u – que foi chamado neste trabalho de ISU^b.

Ambos os métodos são adequados para utilização em planos complexos de amostragem e ambos têm a suposição de que os R24h são não viciados para o consumo usual na escala original. Tem-se também que no método ISU^u os autores assumem que a variabilidade intraindividual pode ser diferente entre os sujeitos, enquanto no ISU^b é pressuposto que essa variabilidade é igual.

O método ISU^u foi proposto por Nusser *et al.* (15) e Guenther *et al.* (16). O organograma que mostra o processo de estimação de forma simplificada foi disposto na Figura 4 a seguir.

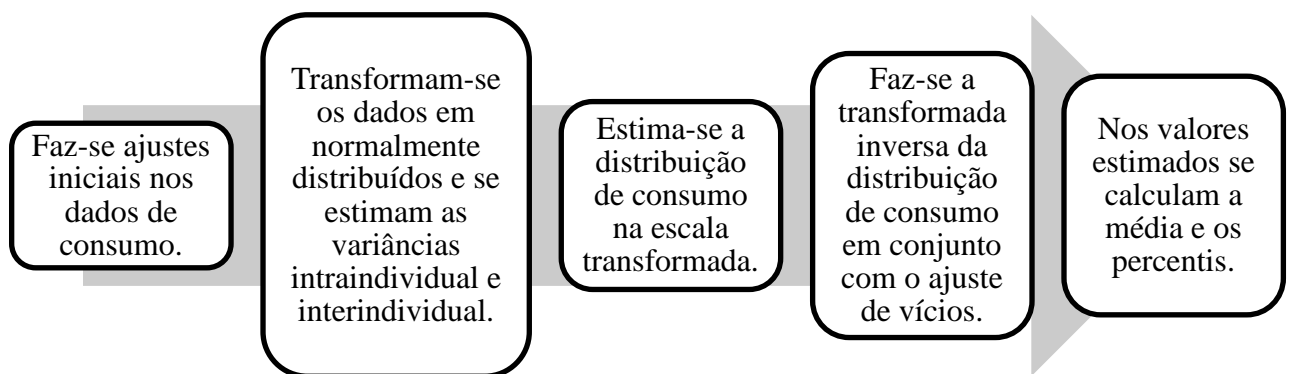


Figura 4. Processo de estimação da distribuição do consumo alimentar usual do método ISU^u.

O método ISU^u consiste nas etapas descritas a seguir:

- Ajustes iniciais dos dados
 - Mudança de dados com consumo zero, somando-se a constante 0,0001 a cada consumo;
 - Ajuste de Regressão para os efeitos perturbadores (mês, dia da semana, tipo de entrevista e sequência da entrevista);

- Transformação inicial em dois estágios, em que, primeiramente, os dados são ajustados para a ocorrência de efeitos perturbadores e, posteriormente, se utiliza a transformação *Power* (17) para normalizar os dados ajustados,
 - Homogeneização dos dias subsequentes à ingestão sobre a distribuição do primeiro dia,
 - Criação de uma amostra com pesos iguais (quando se trata de estudos realizados com amostragem aleatória simples),
 - Transformação inversa para escala original.
- Modelo de Mensuração de erros
 - Transformação total para escala normal (a suavização cúbica *spline* é empregada para estimar a função *Power* que transforma os dados em normais),
 - Estimação das variâncias intraindividual e interindividual (via método dos momentos),
 - Estimação da distribuição de consumo usual na escala transformada,
 - Ajuste dos vícios na transformação para escala original dos valores estimados: a transformação inversa é feita pela estimação da função *Power* via regressão *spline* para a inversa da função utilizada na normalização dos dados em conjunto com o ajuste de vícios.
 - Estimação da distribuição de consumo usual na escala original.

O método ISU^u é implementado no *software* pago que tem duas versões (C-SIDE e PC-SIDE). Ele necessita dos sistemas operacionais UNIX para a versão PC-SIDE e Windows para a versão C-SIDE.

Já o método ISU^b foi descrito por Guenther *et al.* (16) e traz a vantagem de possuir um modelo mais simples e computacionalmente mais rápido que o ISU^u. O organograma que mostra o processo de estimação de forma simplificada foi disposto na Figura 5 a seguir.

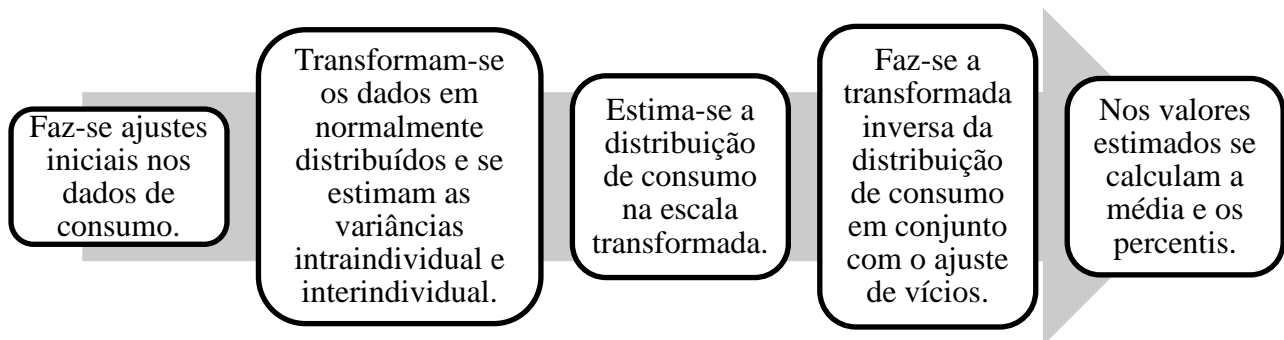


Figura 5. Processo de estimação da distribuição do consumo alimentar usual do método ISU^b.

As etapas que compõe o método ISU^b foram descritas abaixo:

- Ajustes iniciais dos dados
 - Mudança de dados com consumo zero (soma da constante 0,0001 a cada consumo),
 - Transformação inicial em um estágio (transformação *Power* ou *Log*),
 - Ajuste de Regressão para os efeitos perturbadores (mês, dia da semana, tipo de entrevista e sequencia da entrevista),
 - Homogeneização dos dias subsequentes à ingestão sobre a distribuição do primeiro dia,
 - Criação de uma amostra com pesos iguais (quando se trata de estudos realizados com amostragem aleatória simples),
 - Transformação inversa para escala original.
- Modelo de Mensuração de erros

- Transformação total para escala normal (transformação *Power* seguida por uma *Grafted Polynomial*),
 - Estimação das variâncias intraindividual e interindividual (via método dos momentos),
 - Teste de Heterogeneidade para os componentes de variância estimados,
 - Estimação da distribuição de consumo usual na escala transformada,
 - Ajuste dos vícios na transformação para escala original dos valores estimados: a estimação é feita pelo estimador *Grafted Cubic* que utiliza o método dos momentos e a inversa da função que normalizou os dados em conjunto com o ajuste de vícios.
- Estimação da distribuição de consumo usual.

Nusser *et al.* (15) realizaram um estudo de simulação em que compararam os métodos ISU^u e ISU^b e verificaram que o ISU^u tem vícios menores quando comparado ao ISU^b. Contudo, os autores ressaltam que as diferenças entre os dois foram pequenas em termos práticos.

O método ISU^b está implementado em um aplicativo pago que se chama SIDE (*Software for Intake Distribution Estimation*) e ele necessita do módulo SAS/IML para fazer as estimações.

3.1.4. STATISTICAL PROGRAM TO ASSESS DIETARY EXPOSURE (SPADE)

Ele é baseado no programa AGEMODE (28), onde a estimação dos consumos é modelada em função da variável idade e sua interface pode ser dividida em algumas etapas (23).

Antes de detalhar as etapas do método, foi apresentada a Figura 6, na qual está disposto um organograma que mostra de forma simplificada do processo de estimação.

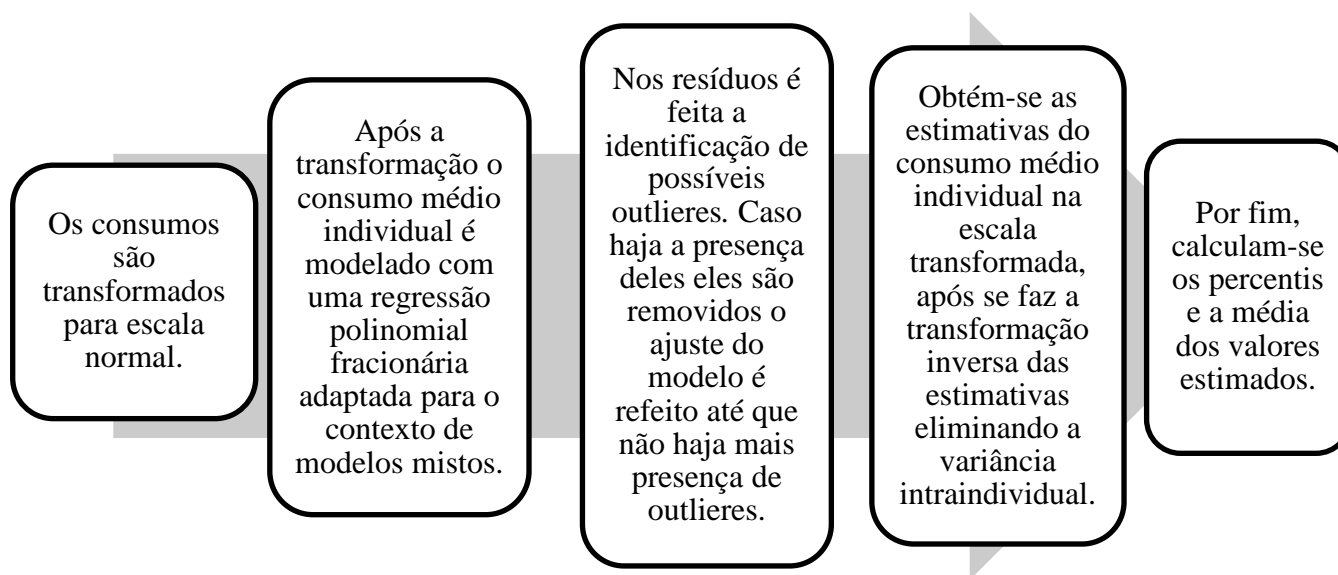


Figura 6. Processo de estimação da distribuição do consumo alimentar usual do método SPADE.

Tem-se que $g(Y_{ij}, \lambda)$, para $Y_{ij} > 0$, é a transformação de Box-Cox de um parâmetro definida como

$$g(Y_{ij}, \lambda) = \begin{cases} \frac{Y_{ij}^{\lambda} - 1}{\lambda} & \text{para } \lambda \neq 0 \\ \log(Y_{ij}) & \text{para } \lambda = 0 \end{cases} \quad (15)$$

onde $g(Y_{ij}, \lambda)$ é a função que transforma os dados em normais e λ é estimado a partir dos dados na escala original.

Após a transformação dos consumos diários individuais, o consumo médio individual é modelado com uma regressão polinomial fracionária. Seja n o número de observações, q e p as potências do polinômio fracionário $Y(x_i)$, tem-se que a função de regressão polinomial é definida como

$$Y_i = a + b(x_i)^p + c(x_i)^q + \varepsilon_i, \text{ para } (i = 1, 2, \dots, n; p \neq q)$$

ou

(16)

$$Y_i^* = a + b(x_i)^p + c(x_i)^q \ln(x_i) + \varepsilon_i, \text{ para } (i = 1, 2, \dots, n; p = q)$$

onde x_i é a idade do indivíduo i e Y_i^* é o consumo transformado do indivíduo i . Já q e p podem assumir valores no intervalo $\{-2; -1; -0,5; 0; 0,5; 1; 2\}$.

Com isso se tem que o consumo transformado é descrito em função da idade por famílias de curvas de no máximo três parâmetros e as frações ideais de q e p são estimadas, assim como a , b e c .

Devido ao fato dos dados de consumo para todos os indivíduos serem para no mínimo dois dias, a equação (16) precisa ser ajustada com um Modelo Misto a fim de se estimar a variância intraindividual (σ_ε^2) e a variância interindividual (σ_u^2). Por isso ela foi redefinida para o caso $p \neq q$ conforme mostra a equação

$$Y_{ij}^* = a + \alpha_i + b(x_i)^p + c(x_i)^q + \varepsilon_{ij} \quad (17)$$

onde y_{ij}^* é o consumo transformado do indivíduo i no dia j , a é o parâmetro aleatório, $\alpha_i \sim N(0, \sigma_u^2)$ e $\varepsilon_{ij} \sim N(0, \sigma_\varepsilon^2)$. Tem-se que os resíduos ε_{ij} são assumidos como normalmente distribuídos com variância constante sobre a idade.

Após o ajuste é feita a identificação de possíveis *outliers* utilizando o teste de qualidade de ajuste de Kolmogorov-Smirnov para os resíduos obtidos na equação (17). Após a retirada dos *outliers*, a transformação Box-Cox é utilizada novamente nos resíduos, uma vez que eles, sem os *outliers*, devem ser simetricamente distribuídos.

Os *outliers* podem influenciar na estimativa da transformada Box-Cox, assim como nos poderes q e p do polinômio fracionário e nas estimativas do Modelo Misto, por isso se houver retirada de *outliers* os passos da estimação devem ser refeitos até que não haja mais presença de *outliers* (os autores relataram que na prática o uso de duas ou três replicações devem ser suficientes).

Em seguida a transformação inversa do consumo médio individual é realizada por Quadratura *Gaussiana* (21). Esse método faz a estimação do consumo médio individual eliminando a variância intraindividual conforme mostra a equação

$$\mu_l = \frac{1}{\sqrt{2\pi\tau^2}} \int_{-\infty}^{+\infty} g^{-1}(v_l + x, \lambda) e^{-\frac{x^2}{2\sigma_\varepsilon^2}} d(x) =$$

$$\frac{1}{\sqrt{\pi}} \int_{-\infty}^{+\infty} g^{-1}(v_{i_l} + \sqrt{2\sigma_\varepsilon^2}y, \lambda) e^{-y^2} d(y) \approx \frac{1}{\sqrt{\pi}} \sum_{j=1}^k w_j g^{-1}(v_l + \sqrt{2\sigma_\varepsilon^2}z_j, \lambda) \quad (18)$$

onde, x é a idade, v_l é a média de consumo do indivíduo l na escala transformada, g^{-1} é a inversa da função da transformada de Box-Cox, já w_j são os pesos e z_j são as abscissas, ou seja, as raízes do polinômio de Hermite para o grau k .

Por fim, a distribuição do consumo alimentar usual é obtida pelo cálculo da média e dos percentis dos valores obtidos na transformação inversa do consumo médio individual.

O SPADE está implementado no *software* R e utiliza os seguintes pacotes: *mpf* (33), *amer* (34), *lme4* (35), *gamlss* (36).

3.1.5. GAY

O método proposto por Gay (18) também é baseado em dados transformados e o organograma que explica de forma simplificada o processo de estimação foi disposto na Figura 7.

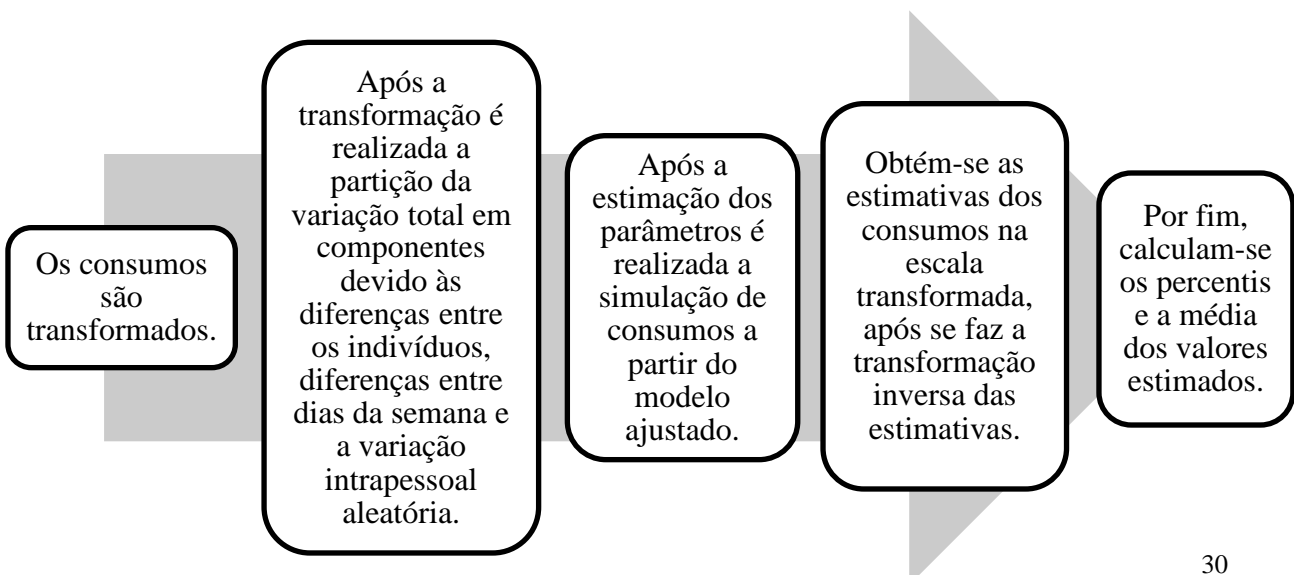


Figura 7. Processo de estimação da distribuição do consumo alimentar usual do método GAY.

Tem-se que a transformação dos dados é realizada pela equação

$$Y_{ij}^* = Y_{ij}^\lambda \quad (19)$$

onde Y_{ij}^* é o consumo transformado do indivíduo i no tempo j , e o valor de λ é determinado por uma rotina automatizada. Nessa rotina a média de consumo e o desvio padrão intraindividual são determinados para cada sujeito para fazer a estimação de λ . Segundo o autor esse valor de λ estabiliza a variabilidade intraindividual e, em muitas vezes, transforma os dados para uma escala aproximadamente normal.

Após a transformação dos dados o método de Análise de Variância (ANOVA, do inglês *Analysis of Variance*) é empregado para realizar a partição da variação total em componentes devido às diferenças entre os indivíduos, diferenças entre dias da semana e a variação intrapessoal aleatória.

Após a estimação dos parâmetros é realizada a simulação de consumos a partir do modelo ajustado, ao passo que a quantidade de dias simulados tem que ser grande o suficiente para eliminar qualquer variação indesejada intraindividual, fazendo com que a estimação seja focada na variabilidade interindividual.

Por fim, os consumos estimados são transformados para escala original e se faz o cálculo das estimativas desejadas.

Esse método utiliza qualquer programa que tenha implementado o método ANOVA e permita a realização de simulação.

3.1.6. SIMPLIFIED NUSSER (S-NUSSER)

Esse método foi proposto no trabalho de Hoffmann *et al.* (19) com o objetivo de ser uma versão mais simples do método ISU. Para facilitar o entendimento foi disposto na Figura 8 um organograma que resume o método de estimação.

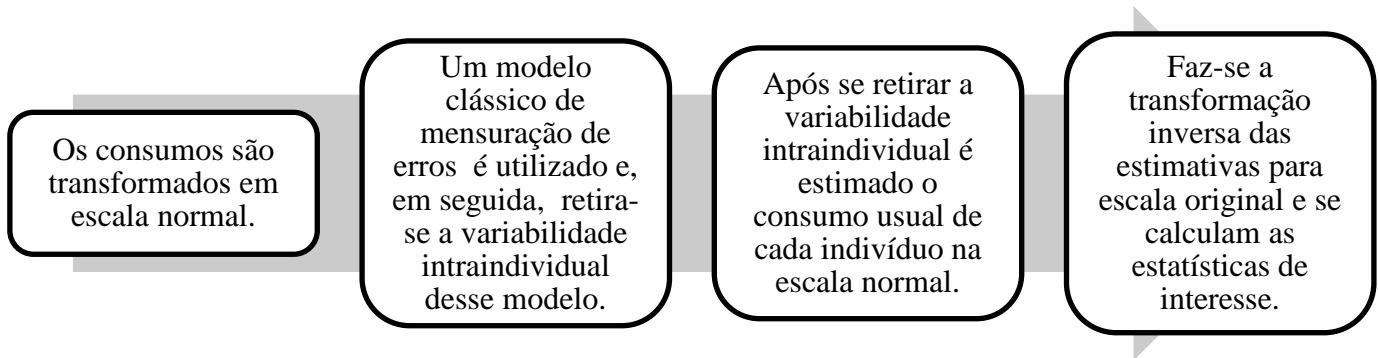


Figura 8. Processo de estimação da distribuição do consumo alimentar usual do método S-NUSSER.

Tem-se que a transformação de Y_{ij} para a normalidade é realizada por meio de uma transformação Box-Cox de dois parâmetros (29):

$$Y_{ij}^* = \begin{cases} ((Y_{ij} + \omega)^\tau - 1)\tau^{-1} & \text{se } \tau \neq 0 \\ \ln(Y_{ij} + \omega) & \text{se } \tau = 0 \end{cases} \quad (20)$$

onde Y_{ij}^* são os consumos transformados na escala normal, os valores de ω e τ são determinados por maximização de Shapiro-Wilk. Para determinar os valores desses parâmetros uma macro do SAS é utilizada e nela τ está definido no intervalo $\left[1, \frac{1}{2}, \frac{1}{3}, \frac{1}{4}, \dots, \frac{1}{10}, 0\right]$ e ω varia ao longo do mesmo intervalo multiplicado pela média dos dados originais.

Após a transformação dos dados é utilizado um modelo clássico de mensuração de erros

$$Y_{ij}^* = T_i + \varepsilon_{ij} \quad (21)$$

onde Y_{ij}^* é o consumo transformado do indivíduo i no dia j , já T_i é referente ao consumo usual verdadeiro do indivíduo i e ε_{ij} é o termo do erro que inclui a variação diária de

consumo e a mensuração aleatória do erro. Os termos T_i e ε_{ij} são supostos independentes entre si e normalmente distribuídos com média igual a zero e variâncias σ_u^2 e σ_ε^2 , respectivamente.

Por conseguinte, tem-se que Y_{ij}^* segue distribuição normal com média μ e variância igual à soma de σ_u^2 e σ_ε^2 , além do mais, a média \overline{Y}_i^* para o indivíduo i - que representa o consumo médio do indivíduo i - tem a mesma distribuição de Y_{ij}^* , mas com variância reduzida conforme mostra a equação a seguir

$$\sigma_{\overline{Y}^*}^2 = \sigma_u^2 + \frac{\sigma_\varepsilon^2}{k} \quad (22)$$

onde k é denotado como o número de dias em que o indivíduo consumiu. Logo, a variável definida por

$$\frac{\sqrt{\frac{\sigma_{\overline{Y}^*}^2 - \sigma_\varepsilon^2}{k}}}{\sigma_{\overline{Y}^*}} [\overline{Y}_i^* - \mu] + \mu \quad (23)$$

tem a mesma distribuição do consumo usual T_i . Portanto, ao se aplicar estimadores padrões para os parâmetros desconhecidos, o consumo usual do indivíduo i na escala normal pode se estimado por

$$\hat{T}_i = \frac{\sqrt{\frac{\hat{\sigma}_{\overline{Y}^*}^2 - \sigma_\varepsilon^2}{k}}}{\hat{\sigma}_{\overline{Y}^*}} [\overline{Y}_i^* - \overline{Y}_{..}^*] + \overline{Y}_{..}^* \quad (24)$$

Já na segunda etapa, o método denota $\hat{\sigma}^2$ como a variância empírica e $\overline{Y}_{..}^*$ é a média geral. Por conseguinte, esse método faz com que se passe de uma média individual \overline{Y}_i^* para uma média geral $\overline{Y}_{..}^*$ e, também, remove a variação intraindividual das médias individuais (σ_ε^2).

Na terceira etapa o consumo usual estimado (\hat{T}_i) é transformado para a escala original utilizando a inversa da função f (vide equação (32)) integrada na distribuição do erro, com ω e τ estimados, conforme mostra a equação

$$f_{retrans}(t) = \lambda^{-\lambda} \sum_{s=0}^{int(\lambda/2)} \binom{\lambda}{2s} (t + \lambda)^{\lambda-2s} \sigma_{\epsilon}^{2s} (2s - 1)!! - \omega \quad (25)$$

onde $f_{retrans}(t)$ é a transformação do consumo usual estimado, $\lambda = \tau^{-1}$, e $x!!$ é denotado como o produto de todos inteiros de 1 a x .

Por fim, nas estimativas na escala original, faz-se o cálculo das estimativas desejadas.

Esse método utiliza o *software* pago SAS nas estimações.

3.1.7. STEM.I E STEM.II

Em 1993 Slob (12) propôs o método STEM.I (*Statistical Exposure Model for Intakes*) para estimação da distribuição do consumo alimentar usual. Já em 2006 ele propôs a extensão do método (13), com a criação do método STEM.II (*Statistical Exposure Model for Incidental Intakes*) que possibilitou tanto a estimação do consumo usual quanto do consumo esporádico.

O método de estimação do consumo usual foi apresentado na sequência e, para facilitar o entendimento, o organograma com o processo de estimação simplificado foi disposto na Figura 9.

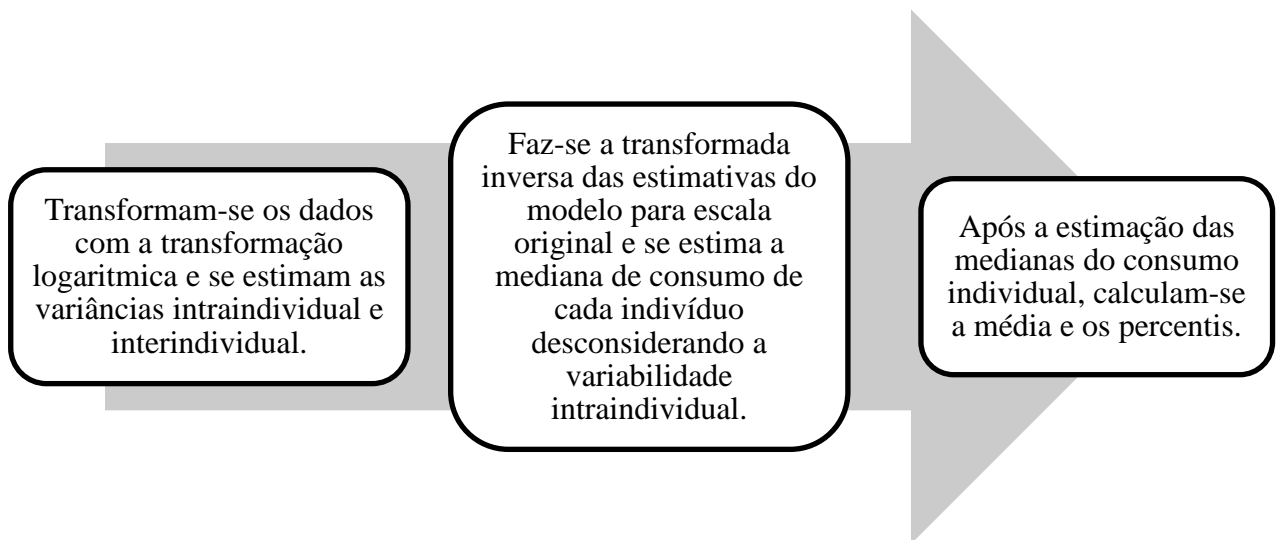


Figura 9. Processo de estimação da distribuição do consumo alimentar usual do método STEM.II.

Seja $Y_{ij}(t)$ referente ao consumo do indivíduo i , com idade t , no dia j , tem-se que o modelo de ajuste desses consumos é descrito pela equação

$$\log[Y_{ij}(t)] = g(\theta_Y; t) + \delta_i + \varepsilon_{ij} \quad (26)$$

onde $\log[Y_{ij}(t)]$ são os dados de consumo transformados para distribuição logarítmica, já $g(\theta_Y; t)$ é o consumo usual para um indivíduo típico, com idade t na escala logarítmica. Tem-se também que δ_i é referente ao desvio do logaritmo do consumo do sujeito i comparado com o geral e ε_{ij} é o desvio do logaritmo do consumo do sujeito i , no dia j comparado ao geral.

O método consiste em estimar o vetor θ_Y dos parâmetros de regressão pelo ajuste da função g para o consumo diário observado em função da idade, ao passo que os resíduos da regressão são obtidos pelo ajuste de uma ANOVA que obtém a estimativa da variância intraindividual (σ_ε^2) que é subtraída da variação total para se obter a variância interindividual (σ_u^2).

O método supõe que no modelo de ajuste:

- Os componentes de variância (σ_u^2 e σ_ε^2) têm distribuição log-normal,
- As variâncias de ε_i e δ_{ij} são homogêneas,
- Que ε_i e δ_{ij} são normalmente distribuídas com média zero.

O autor ressalta que a transformação inversa para escala original de $g(\theta_Y; t)$ reflete a mediana individual na população. Além disso, que o consumo usual - para os dados de consumo diário do indivíduo i - é a mediana (M_i) que tem distribuição lognormal.

Com isso se tem que a mediana (M_i) pode ser estimada pela equação

$$M_i(t) \sim \Lambda(g(t), \sigma_u^2) \quad (27)$$

onde $M_i(t)$ é a mediana de consumo usual do indivíduo i e $\Lambda(g(t), \sigma_u^2)$ é a transformada inversa para escala original das medianas do consumo estimadas pelo

modelo em conjunto com a variância interindividual.

Por fim, a estimação da distribuição do consumo usual é realizada a partir do cálculo da média e dos percentis das medianas do consumo individuais já estimadas.

Esse método utiliza qualquer *software* que realize ANOVA.

3.1.8. WALLACE

Nesse método a distribuição de consumo alimentar usual é estimada por meio de equações que estimam a média geométrica, o desvio padrão geométrico e a média aritmética para as amostras dos dias (14).

O método de estimação do consumo usual foi apresentado na sequência e, para facilitar o entendimento, o organograma com o processo de estimação simplificado foi disposto na Figura 10.

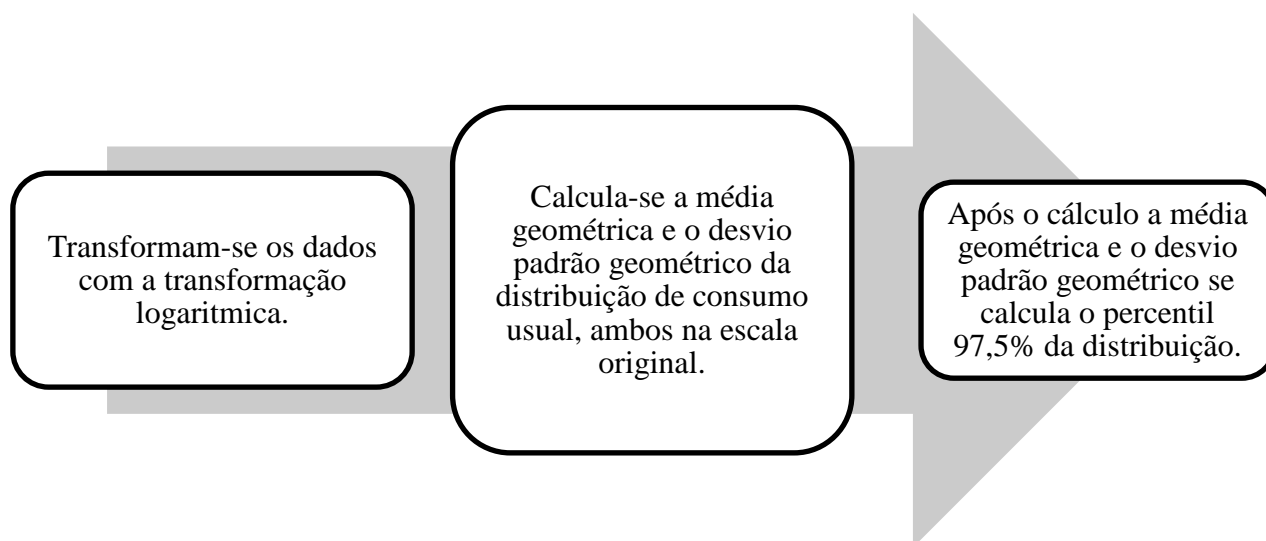


Figura 10. Processo de estimação da distribuição do consumo alimentar usual do método Wallace.

Tem-se que o método assume que o logaritmo deste consumo segue distribuição normal com variância intraindividual σ_{ϵ}^2 e interindividual σ_u^2 , ambas independentes entre si. Com base nessas suposições o desvio padrão geométrico (DPG) é definido como

$$DPG = \exp \left\{ \left[\sigma_u^2 + \ln \left(\frac{(\sigma_\varepsilon^2 - 1)}{r} \right) \right]^{0,5} \right\} \quad (28)$$

onde r é o número de medições do consumo. Tem-se também que a estimação das variâncias é realizada pelo método de mínimos quadrados.

Quando o número de medições é grande ($r \rightarrow \infty$) a variância intraindividual desaparece e o DPG do consumo usual é definido como

$$DPG_U = \exp\{\sigma_u^2\} \quad (30)$$

Tem-se também que a equação para a média geométrica (MG) é definida como

$$MG = \exp \left\{ v + \frac{\sigma_\varepsilon^2}{2} - \frac{1}{2} \ln \left(1 + \frac{(\exp(\sigma_\varepsilon^2) - 1)}{r} \right) \right\} \quad (31)$$

onde v é o logaritmo natural da média geométrica da distribuição observada para uma amostra com um consumo avaliado por pessoa ($r = 1$).

Quando o número de medições é grande ($r \rightarrow \infty$) a MG do consumo usual é definido como

$$MG_U = \exp \left\{ v + \frac{\sigma_\varepsilon^2}{2} \right\} \quad (32)$$

A equação para a média aritmética (MA) é definida como

$$MA = \exp \left\{ v + \frac{(\sigma_\varepsilon^2 + \sigma_u^2)}{2} \right\} \quad (33)$$

Tem-se também que o percentil 97,5% é definido pelo produto de MG_U e DPG_U conforme mostra a equação a seguir

$$Percentil\ 97,5\% = MG_U \times (DPG_U)^2 \quad (34)$$

Após o cálculo da média geométrica e do desvio padrão geométrico se calcula o percentil 97,5% da distribuição de consumo alimentar usual.

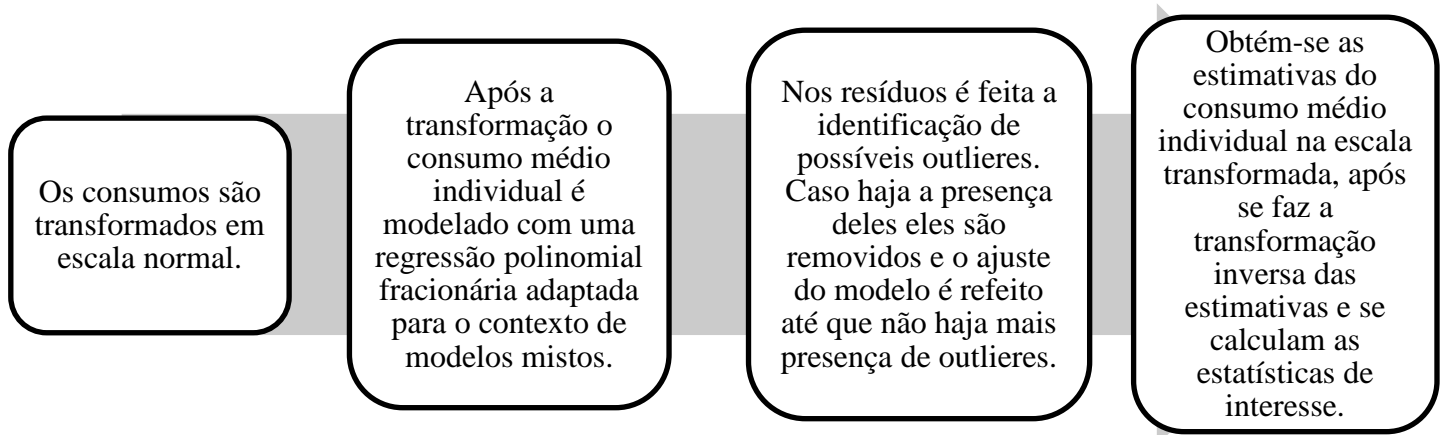
Esse método está implementado no software *Small*.

3.1.9. AGEMODE

O AGEMODE foi proposto por Waijers *et al.* (28) e nesse método a estimação

dos consumos é modelada em função da variável idade. Ele é similar ao método SPADE, apresentado anteriormente, pois o a metodologia de estimação dos dois é semelhante diferindo apenas na forma de realizar a transformação inversa das estimativas para escala original.

Para facilitar o entendimento foi disposto na Figura 11 um organograma que



resume o método de estimação.

Figura 11. Processo de estimação da distribuição do consumo alimentar usual do método AGEMODE.

A primeira parte da modelagem transforma os consumos em escala normal via uma transformação Box-Cox de um parâmetro (29).

Tem-se que $g(Y_{ij}, \lambda)$ é a transformação de Box-Cox de um parâmetro definida como

$$g(Y_{ij}, \lambda) = \begin{cases} \frac{Y_{ij}^\lambda - 1}{\lambda} & \text{para } \lambda \neq 0 \\ \log(Y_{ij}) & \text{para } \lambda = 0 \end{cases} \quad (35)$$

onde, $g(Y_{ij}, \lambda)$ é a função que transforma os dados para escala normal e λ é o valor utilizado pela transformada de Box-Cox na transformação dos dados que é estimada a partir deles. Após a transformação o consumo médio individual é modelado por uma

regressão polinomial fracionária.

Seja n o número de observações e q e p as potências do polinômio fracionário $Y(x_i)$, tem-se que a função de regressão polinomial é definida como

$$Y_i^* = a + b(x_i)^p + c(x_i)^q + \varepsilon_i, \text{ para } (i = 1, 2, \dots, n; p \neq q)$$

ou (36)

$$Y_i^* = a + b(x_i)^p + c(x_i)^q \ln(x_i) + \varepsilon_i, \text{ para } (i = 1, 2, \dots, n; p = q)$$

onde x_i é a idade do indivíduo i e Y_i^* é o consumo transformado do indivíduo i ; já q e p podem assumir valores no intervalo $\{-2; -1; -0,5; 0; 0,5; 1; 2\}$.

Com isso se tem que o consumo transformado é descrito em função da idade por famílias de curvas de no máximo três parâmetros, ao passo que as frações ideais de q e p são estimadas assim como a , b e c .

Devido ao fato dos dados de consumo para todas as pessoas serem para no mínimo dois dias, a equação (36) precisa ser ajustada com um Modelo Misto, a fim de se estimar a variância intraindividual (σ_ε^2) e a variância interindividual (σ_u^2). Por isso, os autores redefinem a equação (36) para o caso $p \neq q$ definindo a como parâmetro aleatório:

$$Y_{ij}^* = a + \alpha_i + b(x_i)^p + c(x_i)^q + \varepsilon_{ij} \quad (37)$$

onde Y_{ij}^* é o consumo transformado do indivíduo i no dia j , $\alpha_i \sim N(0, \sigma_u^2)$ e $\varepsilon_{ij} \sim N(0, \sigma_\varepsilon^2)$, tem-se também que os resíduos ε_{ij} são assumidos como normalmente distribuídos com variância constante sobre a idade.

Após o ajuste é feita a identificação de possíveis *outliers* utilizando o teste de qualidade de ajuste de Kolmogorov-Smirnov para os resíduos obtidos na equação (37). Após a retirada dos *outliers* a transformação Box-Cox é utilizada novamente nos

resíduos, uma vez que eles, sem os *outliers*, devem ser simetricamente distribuídos.

Os *outliers* podem influenciar na estimativa da transformada Box-Cox, assim como nos poderes q e p do polinômio fracionário e nas estimativas do Modelo Misto, por isso se houver retirada de *outliers* os passos da estimação devem ser refeitos até que não haja mais presença de *outliers* (os autores relatam que na pratica o uso de duas ou três replicações devem ser suficiente).

Em seguida, são gerados consumos na escala transformada - via simulação Monte Carlo - com base nos resultados obtidos no Modelo Misto. Com isso, geram-se consumos para n sujeitos em cada idade e, ainda, cria-se uma série temporal de k consumos, para cada sujeito i com $\alpha \sim N(0, \sigma_u^2)$ e $\varepsilon_{i,t} \sim N(0, \sigma_\varepsilon^2)$ onde $t = 1, 2, \dots, k$. A série temporal para cada indivíduo i é definida como

$$Y_{i,t}^* = a + b(x_i)^p + c(x_i)^q + d_i + e_{i,t} \quad (38)$$

onde d_i e $e_{i,t}$ são realizações de α_i e ε_{ij} na equação (37).

Em seguida em cada uma das observações geradas é realizada a transformação inversa para escala original pela equação abaixo

$$Y_{i,t} = (Y_{i,t}^* + 1)^{1/\lambda} \quad (39)$$

Após a transformação inversa dos dados na escala original são calculados os percentis e a média.

Esse método utiliza o *software* pago S-PLUS nas estimações.

3.2. COMPARAÇÃO DE MÉTODOS

Alguns autores realizaram estudos onde compararam os resultados de diversos métodos de estimação da distribuição de consumo alimentar usual, mas a falta de uma distribuição teórica para a distribuição fez com que os estudiosos dos métodos aproximassem os verdadeiros valores a serem comparados a partir de técnicas distintas,

com o objetivo de verificar a precisão dos métodos.

Dodd *et al.* (11) utilizaram biomarcadores para estimar a verdadeira densidade do consumo de nutrientes, enquanto Guenther *et al.* (16) usaram a densidade estimada dos dados - obtida no estudo USDA entre 1989-1981 - para estimar a densidade da distribuição usual verdadeira, por outro lado Hoffmann *et al.* (19) empregaram como meio de comparação as estimativas obtidas nas repetições dos diversos R24h utilizados no estudo EPIC-Potsdam.

Por outro lado Nusser *et al.* (15) utilizaram a integração numérica para estimar a verdadeira distribuição do consumo usual dos dados gerados por eles no seu estudo de simulação, já Tooze *et al.* (27) usaram a simulação da média de consumo para 365 dias de 250.000 indivíduos simulados, ao passo que Goedhart *et al.* (22) empregaram uma técnica parecida onde estimaram a verdadeira distribuição de consumo usual via simulação do consumo de dois dias de 100.000 indivíduos. Também há o trabalho de Souverein *et al.* (23) no qual realizada a estimação dos verdadeiros valores para a média e os percentis dos consumos via aproximação por *Quadratura Gaussiana*, fazendo uso dos parâmetros utilizados na geração dos dados simulados.

Todas as técnicas de comparação citadas são válidas, mas não mostram necessariamente os valores verdadeiros para a distribuição. No caso dos biomarcadores, eles podem não mostrar exatamente o quanto do nutriente os indivíduos consomem, pois a medição depende do mecanismo de digestão que difere de uma pessoa para outra, além do fato de só poderem ser avaliados para o consumo de nutrientes. Já a utilização de estimação pela *Quadratura Gaussiana* também pode gerar flutuações, pois esse método é uma forma de estimação da distribuição, o que também vale para os casos em que se estimou a verdadeira distribuição por meio de simulação, observação de vários R24h e estimação da densidade empírica a partir de grandes bases de dados.

Apesar de haver vários trabalhos que fazem comparações dos métodos, destaca-se o trabalho de Souverein *et al.* (23) que foi um dos poucos que realizou simulação de dados e comparou um número razoável de métodos. Nesse trabalho se realizou um estudo de simulação comparando quatro métodos para estimar a distribuição do consumo alimentar usual (ISU, MSM, NCI e SPADE) com o objetivo de averiguar o quanto as distribuições estimadas de consumo alimentar usual desviam da verdadeira distribuição.

Nesse trabalho os autores definiram os seguintes parâmetros para a simulação dos dados: a média de consumo geral (μ), o desvio-padrão interindividual (σ_u), o desvio-padrão intraindividual (σ_ε), a razão de variâncias definida pela razão das variâncias interindividual e intraindividual ($r_{var} = (\sigma_\varepsilon^2)/(\sigma_u^2)$), o lambda da transformação inversa de Box-Cox (λ) - onde $\lambda = 0$ é a distribuição Log-Normal e $\lambda = 1$ é equivalente à distribuição Normal - e o tamanho amostral (n).

Para simular o consumo alimentar usual na escala transformada de um indivíduo - para cada um dos n sujeitos - utilizou-se uma distribuição normal com média μ e desvio-padrão interindividual. Já quando se referia ao consumo de dois dias na escala transformada o consumo médio de cada dia foi obtido a partir de uma distribuição normal com média igual à gerada para o indivíduo - obtidas no passo anterior - e desvio-padrão intraindividual.

Para avaliar as estimativas da distribuição do consumo alimentar usual foram simulados dez diferentes cenários, com 100 replicações cada: os três primeiros foram concebidos para comparar o efeito de tamanho de amostra, já o quarto, quinto e sexto cenários para checar a influência da assimetria e, por fim, os últimos foram utilizados para conferir o efeito da razão entre as variâncias.

As gerações dos cenários foram realizadas com as definições a seguir: para os

três primeiros cenários foram utilizados, respectivamente, os seguintes tamanhos amostrais: 150, 500 e 1000, mantendo-se constantes os valores de $\lambda = 0$ e $r_{var} = 1$. Já para o quarto, quinto e sexto cenários os valores de λ utilizados foram 0,0; 0,2 e 0,5, mantendo-se constantes os valores de $n = 1000$ e $r_{var} = 1$. Por fim, os últimos quatro cenários apresentaram os valores da r_{var} iguais a 0,25; 1; 4 e 9, mantendo-se constante os valores de $n = 1000$ e $\lambda = 0$.

Após a geração dos dados de todos os cenários foram aplicados os métodos de estimação de consumo alimentar usual, em cada amostra de cada cenário, e foram calculados os percentis e a média.

Contudo, houve limitações neste estudo, como, por exemplo, a utilização de 100 replicações para cada cenário, uma vez que geralmente estudos de simulação trabalham com 1.000 ou 10.000 replicações para poder dar estabilidade às estimativas.

Outras limitações foram o enfoque do estudo em dados com distribuição Log-normal e a estimação dos percentis verdadeiros utilizando o método de estimação da Quadratura *Gaussiana*, pois esse método realiza apenas uma aproximação dos percentis verdadeiros.

4. OBJETIVO

Objetivo Geral:

- Propor o novo método de estimação da distribuição do consumo alimentar usual
- MEID - e compará-lo com outros já existentes.

5. ARTIGO

A ser enviado ao European Journal of Clinical Nutrition.

MEID: um novo método para estimação da distribuição de consumo alimentar usual e comparação com NCI, SPADE e MSM.

MEID: a new method for estimating the distribution of usual food intake and comparison with NCI, SPADE and MSM.

G H C Laureano¹; S A Camey^{1,2} e V B L Torman²

UNIVERSIDADE FEDERAL DO RIO GRANDE DO SUL (UFRGS)

CONHECIMENTO/OBJETIVOS: Já há inúmeros métodos de estimação da distribuição do consumo alimentar usual. De uma forma geral, eles são de difícil entendimento e implementados em *softwares* pagos. Em vista disso, este trabalho propõe o novo método de estimação da distribuição do consumo alimentar usual intitulado como MEID (*Method for Estimate Intake Distribution*). Este método tem como objetivo ser mais simples que os já existentes, uma vez que em sua metodologia não se utiliza a transformação inversa para escala original das estimativas.

SUJEITOS/MÉTODOS: Neste trabalho este novo método foi comparado com os métodos SPADE (*Statistical Program to Assess Dietary Exposure*), NCI (*National Cancer Institute*) e MSM (*Multiple Source Method*) por meio de um estudo de simulação. Para cada método foram estimados média e percentis da distribuição do consumo usual e foram utilizadas como medidas de acurácia o vício e o vício relativo, já como medida de precisão foi utilizado o Erro Quadrático Médio (EQM). O método proposto neste trabalho utilizou o *software* livre R em todo seu escopo.

RESULTADOS: Os métodos NCI, SPADE e MSM realizaram estimações parecidas e obtiveram EQM e vícios superiores ao MEID, principalmente quando se tratava do percentil 95.

CONCLUSÕES: O estimador MEID foi mais acurado e preciso na estimação da distribuição do consumo alimentar usual quando comparado aos demais, como também se mostrou mais simples e de interface implementável em qualquer *software* com linguagem similar ao do *software* R.

Palavras chave: Estimação, comparação, SPADE, NCI e MSM.

KNOWLEDGE / OBJECTIVES: There are already numerous methods for estimating the distribution of usual dietary intake. In general, they are difficult to understand and implemented in paid software. In this light, this paper proposes a new method for estimating the distribution of usual dietary intake: MEID (Method for Intake Distribution Estimate). This method aims to be simpler than existing ones, since in their methodology not use the back-transformation to the original scale of the estimates.

SUBJECTS / METHODS: In this work this method was compared with the SPADE (Statistical Program to Assess Dietary Exposure), NCI (National Cancer Institute) and MSM (Multiple Source Method) methods, by means of a simulation study. For each method, the mean and the percentiles of usual intake were estimated, and for accuracy

¹ Programa de Pós-graduação em Epidemiologia, UFRGS - greice.laureano@ufrgs.br 45

² Departamento de Estatística - UFRGS.

measurements, the bias and relative bias were used. The precision was measured using the Mean Squared Error (MSE). The proposed method in this paper uses the free software R.

RESULTS: The results have shown that the NCI, MSM and SPADE methods performed similar estimations, and obtained MSE and bias higher than MEID, especially in the 95th percentile.

CONCLUSIONS: In short, it was found that the MEID estimator was more accurate and precise in estimating the distribution of usual dietary intake, when compared to other estimators, while also being simpler than those. It was also shown that it's implementable in any language similar to the R software.

Keywords: Estimation, comparison, MEID, SPADE, NCI e MSM

INTRODUÇÃO

A mensuração do consumo alimentar usual é assunto vigente na área da saúde, uma vez que várias doenças são influenciadas, ou até mesmo causadas, pelos hábitos alimentares do indivíduo (1). Uma das formas de estudar o consumo alimentar de uma população é por meio da estimação da distribuição do consumo usual. O estudo dessas distribuições traz um melhor entendimento a respeito do consumo alimentar em populações, como também, pode ajudar na identificação de possíveis grupos de risco para o desenvolvimento de doenças.

Os métodos de estimação da distribuição do consumo alimentar usual usam as informações de estudos que avaliam o consumo de no mínimo dois dias por sujeito observado. Salienta-se que não há um método de mensuração do consumo alimentar que seja definido como padrão-ouro na área de nutrição, uma vez que existem diversos métodos – como, por exemplo, os Recordatórios Alimentares de 24 horas (R24h), os Questionários de Frequência Alimentar (QFA), os registros alimentares e a pesagem de alimentos.

Para estimar a distribuição de consumo alimentar usual foram propostas diversas metodologias (2–17), mas elas, de uma forma geral, são de difícil entendimento devido à complexa metodologia estatística envolvida e também são limitadas ao uso de *softwares* específicos e/ou pagos. Em vista disso, este trabalho propõe um novo método de estimação da distribuição do consumo alimentar usual: MEID (*Method for Estimate Intake Distribution*). A proposição desse novo estimador tem como objetivo ser de fácil entendimento, aplicável ao *software* livre R (18) com

uma metodologia mais simples que métodos já existentes.

Alguns autores realizaram estudos onde comparavam métodos de estimação da distribuição do consumo alimentar usual (6–8,14,19–21). Entretanto a falta de uma distribuição teórica, ou seja de um padrão ouro, para a distribuição de consumo alimentar usual fez com que os estudiosos dos métodos tivessem que arbitrar sobre a distribuição teórica, para poderem avaliar a qualidade dos métodos.

Souverein *et al.* (19) utilizaram a estimação dos verdadeiros valores para a média e os percentis dos consumos via aproximação por *Quadratura Gaussiana*, fazendo uso dos parâmetros utilizados na geração dos dados simulados por eles. Já Dodd *et al.* (20) utilizaram biomarcadores para estimar a verdadeira densidade do consumo de nutrientes, enquanto Guenther *et al.* (7) utilizaram a densidade estimada dos dados obtida a partir dos dados do estudo USDA entre 1989-1981. Por outro lado Hoffmann *et al.* (8) utilizaram como meio de comparação as estimativas obtidas nas repetições dos diversos R24h utilizados no estudo EPIC-Potsdam. Já Nusser *et al.* (6) usaram integração numérica para estimar a verdadeira distribuição do consumo usual dos dados gerados por eles no seu estudo de simulação. Tooze *et al.* (14) fizeram uso da simulação da média de consumo para 365 dias de 250.000 indivíduos simulados, ao passo que Goedhart *et al.* (21) valeram-se de uma técnica parecida onde estimaram a verdadeira distribuição de consumo usual via simulação do consumo de dois dias de 100.000 indivíduos.

Todas as técnicas citadas no parágrafo anterior são válidas, mas não mostram necessariamente os valores

verdadeiros para a distribuição de consumo alimentar usual. No caso dos biomarcadores, eles podem não mostrar exatamente o quanto do nutriente os indivíduos consomem, pois a medição depende do mecanismo de digestão que difere de uma pessoa para outra. Da mesma forma, a utilização de estimação pela Quadratura *Gaussiana* também pode gerar flutuações, pois esse método é uma forma de estimação da distribuição. Isto também vale para os casos em que se estimou a verdadeira distribuição por meio de simulação, observação de vários R24h e estimação da densidade empírica a partir de grandes bases de dados.

Dentre os estudos que realizaram comparações dos métodos se destaca o trabalho de Souverein *et al.* (19) que foi um dos poucos que realizou simulação de dados e comparou um número razoável de métodos. Eles compararam os métodos MSM, NCI, SPADE e ISU (*Iowa State University Method*) avaliando a influência na qualidade da estimação dos métodos quando se têm tamanhos de amostras distintos, diferenças entre as variâncias intraindividual e interindividual e diferentes valores para o parâmetro da transformada de Box-Cox (22).

Contudo, houve limitações neste estudo, como, por exemplo, a utilização de apenas 100 replicações para cada cenário. Outras limitações foram o enfoque do estudo em dados com distribuição Lognormal e o já mencionado cálculo dos percentis verdadeiros utilizando o método de estimação da Quadratura Gaussiana.

Neste trabalho realizamos um estudo de simulação com a mesma abordagem de Souverein *et al.* (19), onde comparamos os métodos SPADE, NCI e MSM com o novo método de estimação proposto. Um número maior de replicações foi utilizado e deu-se enfoque aos cenários que apresentaram maior incerteza nas conclusões. Além disto, foi possível obter os verdadeiros valores dos parâmetros a partir da densidade da distribuição Box-Cox.

Ressalta-se que o método MEID não incorpora a possibilidade de dados com consumo esporádico nem o uso de covariáveis. O seu desenvolvimento também considerou a estimação do consumo de apenas um alimento ou nutriente no contexto de amostragem

aleatória simples.

Maiores detalhes a respeito do MEID serão apresentados na próxima sessão, como também serão ilustrados os demais métodos utilizados neste estudo e como foi realizada a simulação. Logo em seguida, serão apresentados os resultados, a discussão e as conclusões deste trabalho.

MÉTODOS

O consumo alimentar usual, em geral, tem distribuição assimétrica (20), o que impossibilita a utilização de métodos estatísticos baseados na distribuição Normal. Nesses casos uma das táticas utilizadas é a transformação dos dados.

Quase todos os métodos existentes (2–17) estimam as variâncias (intraindividual e interindividual), e a média do consumo da população, na escala transformada. Posteriormente a obtenção dessas estimativas (na escala normal), faz-se a transformação inversa para a escala original e então se obtém a média e os percentis da distribuição do consumo alimentar usual.

Esses métodos diferem em relação às formas de transformação, aos modelos para o ajuste dos dados, às formas de se obter a transformação inversa e aos métodos de obtenção da média e dos percentis da distribuição.

O método MEID trabalha com a transformação do consumo para distribuição normal através da transformação Box-Cox de um parâmetro (22).

Para descrever a metodologia utilizada, seja y_{ij} a quantidade consumida de alimento, ou nutriente, pelo sujeito i no momento j . Assim, nesse caso, a transformação Box-Cox de um parâmetro é definida da seguinte forma:

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

onde y_{ij}^* é o consumo do sujeito i no momento j transformado para a escala normal e λ é o parâmetro da distribuição Box-Cox.

O parâmetro λ é estimado a partir dos dados de consumo pelo pacote *car* (23) do *software* R.

Com isso, após a transformação dos dados, um Modelo Misto (24) é ajustado nos valores transformados, com

o objetivo de separar a variabilidade interindividual e intraindividual, para tanto se utiliza o pacote *lme4* (25) do *software* R. Temos que o Modelo Misto é descrito como:

$$y_{ij}^* = \mu + u_i + \varepsilon_{ij}; \quad i = 1, 2, \dots, n; \quad j = 1, 2, \dots, r \quad (2)$$

onde n é o número de sujeitos, r é o número de vezes que foi observado o R24h ($r > 2$), μ é a média do consumo transformado da população, u_i é o efeito aleatório que é pressuposto ter distribuição $N(0, \sigma_u^2)$, já o ε_{ij} é o erro de mensuração relativo ao sujeito i no momento j que é pressuposto serem independente de u_i e ter distribuição $N(0, \sigma_\varepsilon^2)$. Isso implica que

$$E[y_{ij}^*] = \mu \quad (3)$$

e que a variância pode ser decomposta da seguinte forma:

$$Var[y_{ij}^*] = \sigma_u^2 + \sigma_\varepsilon^2 \quad (4)$$

onde σ_u^2 é a variância interindividual e σ_ε^2 é a variância intraindividual.

O Modelo Misto citado fornece estimativas de μ e σ_ε^2 , utilizadas para calcular a média e os percentis da densidade teórica da distribuição Box-Cox de um parâmetro (22). Esses parâmetros foram calculados pelo pacote *gamlss.dist* do *software* R.

O método MEID neste trabalho foi implementado no *software* livre R versão 3.0.1.

O MSM (15,27) está implementado na web no site: <http://msm.dife.de/>.

O NCI foi proposto por Tooze et al. (9,10,13,14). Ele está implementado no SAS (28) e pode ser encontrado no site: <http://riskfactor.cancer.gov/diet/usualintakes/>.

O SPADE (17,19,21) está implementado no *software* R e é baseado no programa AGEMODE (12) onde a estimação dos consumos é modelada em função da covariável idade. Entretanto, mesmo o SPADE sendo modelado em função da idade, essa informação pode não ser fornecida mediante alguns ajustes no código do método.

O Quadro 1 mostra de forma resumida as características do método proposto e dos métodos MSM, NCI e SPADE quando se referem a estimação da distribuição do consumo alimentar usual.

Apesar deste trabalho ter seguido a metodologia utilizada por Souverein et

al. (19) não foi feita a comparação com o método ISU, pois sua interface impossibilitou o aumento de replicações da simulação, uma vez que além de ser um *software* pago a seleção de menus do mesmo é feita de forma manual.

A simulação dos dados foi definida para o consumo de um alimento cuja distribuição seja Box-Cox. Para tanto foram definidos os seguintes parâmetros: a média de consumo geral (μ), o desvio-padrão interindividual (σ_u), o desvio-padrão intraindividual (σ_ε), a razão de variâncias intraindividual interindividual ($r_{var} = (\sigma_\varepsilon^2/\sigma_u^2)$) - ambos na escala normal - e o parâmetro da transformação inversa de Box-Cox (λ). Com os valores para os parâmetros definidos os dados de consumo foram gerados da seguinte forma:

- Primeiramente, foi gerado o consumo médio individual dos n sujeitos a partir de uma distribuição normal com média $\mu = 7,5$ (que gera um consumo médio na escala original de aproximadamente 100 gramas) e desvio padrão interindividual (σ_u),
- Posteriormente, foram geradas duas observações de consumo por sujeito (isto é $r = 2$), utilizando o consumo médio individual do passo anterior e o desvio padrão intraindividual (σ_ε),
- Por fim, foi utilizada a transformação inversa de Box-Cox (λ) para transformar os dados na escala original conforme a equação:

$$y_{ij} = \sqrt[\lambda]{\lambda y_{ir}^* + 1};$$

$$i = 1, 2, \dots, n; \quad r = 2. \quad (5)$$

onde, y_{ij} são os dados na escala original, λ é o valor estimado da transformação Box-Cox, n é igual ao número de sujeitos e k é igual ao número de consumos simulados por sujeito.

Foram simulados dados para quatro cenários diferentes, pois nesses cenários os resultados encontrados por Souverein et al. (19) apresentaram maior variabilidade e, portanto, incerteza nas conclusões. Os cenários foram nomeados da seguinte forma: Cenário I ($n = 500$ e $r_{var} = 4$), Cenário II ($n = 500$ e $r_{var} = 9$), Cenário III ($n = 150$ e $r_{var} = 4$) e Cenário IV ($n = 150$ e $r_{var} = 9$). Para cada cenário foram utilizadas 1.000 replicações.

A variância intraindividual foi fixada ($\sigma_{\varepsilon}^2 = 1$) para todos os cenários e se variou o valor da variância interindividual de acordo com o valor da razão de variância a ser estudada.

Na simulação foram utilizados os softwares R (para a geração dos dados e os ajustes dos métodos MEID e SPADE), SAS (para o ajuste do NCI) e o AutoHotkey (29) (para o simulação do MSM).

Para comparação das estimativas se calculou o vício, o vício relativo e o EQM para cada método. Para o cálculo dessas medidas foi necessário obter a média e os percentis verdadeiros e, para isso, utilizou-se a densidade da distribuição Box-Cox implementada no pacote *gamlss.dist* (26) do *software* R. Esse pacote permite obter a densidade a partir da informação dos parâmetros μ , variância entressujeito (ambos na escala normal) e λ . A partir dessa densidade é possível obter os percentis e a média da distribuição verdadeira.

O código para a geração de dados e estimação do método MEID foi disposto no material suplementar deste artigo e pode ser aplicado em qualquer banco de dados que contenha no mínimo dois dias de consumo por indivíduo.

RESULTADOS

A Tabela 1 mostra os verdadeiros valores, a média e o vício relativo das estimativas da média e dos percentis da distribuição do consumo alimentar usual para cada cenário. Nessa tabela é possível observar que o método MEID obteve as estimativas em média mais próximas dos valores verdadeiros. Isto implicou que o estimador MEID obteve os menores vícios relativos.

Na Tabela 2 são apresentados os EQMs para cada estimador em cada cenário. Nela é possível verificar que o estimador MEID obteve os menores EQMs quando comparado com os demais.

Na Figura 1 são apresentados os *Box-Plot's* para os vícios dos estimadores em cada cenário. Nesses gráficos se verifica que o estimador MEID tem as menores dispersões das estimativas quando comparado aos demais. Entretanto, é evidente que o método teve resultados prejudicados na estimação do percentil 95 - em todos os cenários - como também em todas estimativas dos

cenários 3 e 4.

O método NCI não conseguiu fazer as estimações para alguns bancos de dados gerados, assim como o método MEID. O Quadro 2 mostra a quantidade de simulações perdidas para cada método em cada cenário. A ocorrência desses problemas na estimação se deu ao fato dos métodos algumas vezes estimarem variâncias negativas ou nulas para a variância interindividual e, por conseguinte, não conseguiram finalizar as estimações.

DISCUSSÃO/ CONCLUSÕES

Neste trabalho foi proposta uma nova metodologia para estimação da distribuição de consumo alimentar usual que tinha como objetivo ser mais simples e que pudesse ser utilizada no *software* livre R. O objetivo foi atingido e ainda obteve-se um método no mínimo tão preciso quanto os já existentes.

Os resultados obtidos nos cenários simulados mostraram que em termos de acurácia - o tamanho de amostra interferiu na qualidade das estimações de todos métodos. Quando se trata da precisão é possível verificar que o tamanho de amostra também influencia em todos os métodos. Já o aumento da razão de variâncias provocou um decréscimo da acurácia e da precisão de todos os métodos.

Os resultados também mostraram que há uma estimação de pior qualidade em todos os métodos quando se trata da estimação dos percentis 5 e 95, sendo o último o pior caso. Como o maior interesse na estimação da distribuição alimentar em populações é avaliar se a poluição tem deficiências alimentares ou excessos alimentares a estimação desses percentis é de suma importância, por isso é necessário que os métodos tenham alta precisão e acurácia na estimação desses percentis.

Em suma o MEID obteve as estimações mais precisas e acuradas em todos cenários. Tem-se também que ele apresenta uma metodologia mais simples que as dos demais métodos, uma vez que não utiliza a transformação inversa para escala original. Ele também pode ser implementado em qualquer *software* que tenha linguagem compatível com a do *software* R.

A extensão do método para outras

famílias de distribuições diferentes da Box-Cox se faz necessária para abranger dados que possam não se ajustar bem a esse modelo. Também ainda é necessário, mas viável, estender o método MEID para a inserção de covariáveis (os métodos MSM, NCI e SPADE permitem essa inserção).

Ressalta-se que apesar das limitações, para estimação de distribuição alimentar usual, sem covariáveis, o método MEID já pode ser plenamente aplicado e é recomendado por ser mais acurado e preciso.

Quadro 1. Resumo da metodologia dos métodos comparados

Etapa	Método			
	MSM	NCI	SPADE	MEID
Transformação	Os resíduos do modelo ajustado são transformados para uma distribuição normal pela transformada de Box-Cox de dois parâmetros.	Transforma os consumos observados em normais com uma transformação Box-Cox de um parâmetro.		
Modelagem	É utilizado um modelo de regressão linear múltipla no ajuste dos dados de consumo, por conseguinte os resíduos são transformados em normais e se estima a média e as variâncias intraindividual e interindividual.	É utilizado um Modelo Misto Não Linear nos dados de consumo transformados e se estima a média e as variâncias intraindividual e interindividual.	O consumo médio individual é modelado por uma regressão polinomial fracionária e as variâncias intraindividual e interindividual são estimadas.	Nos dados transformados se estima a média e as variâncias intraindividual e interindividual utilizando Modelo Misto.
Simulação	Não utiliza.	Simulam-se k (valor a ser fixado) pseudos consumos com distribuição normal e desvio padrão igual ao interindividual para cada sujeito.	Não utiliza.	Não utiliza.
Transformada Inversa	É definida como a transformação da função dos resíduos normais integrada na distribuição do erro.	É feita no conjunto de dados simulados pela aproximação via série de Taylor.	A transformação inversa do consumo médio individual é realizada pelo método de <i>Quadratura Gaussiana</i> .	Não utiliza.

Etapa	Método			
	MSM	NCI	SPADE	MEID
Cálculo das médias e percentis	Nos consumos estimados pelo modelo de regressão, após a transformação inversa para escala original dos resíduos.	A partir dos dados simulados transformados para a escala original, via interpolação linear.	A partir do consumo médio individual transformado para escala original.	A partir das estimativas de λ , μ e σ_ε^2 para obtê-las diretamente da densidade da distribuição Box-Cox de um parâmetro.
Abreviações: NCI, <i>National Cancer Institute</i> , MSM, <i>Multiple Source Method</i> , SPADE, <i>Statistical Program to Assess Dietary Exposure</i> , MEID, <i>Method for Estimate Intake Distribution</i> .				

Quadro 2. Resumo das simulações perdidas por cenário e método.

Método	Cenário			
	I	II	III	IV
MEID	0	14	12	119
NCI	0	1	1	12

Tabela 1. Médias e vícios relativos das estimativas obtidas pelos estimadores para os cenários.

Cenário	Parâmetro	Valor Verdadeiro	Média (Vício Relativo %)							
			NCI		MSM		SPADE		MEID	
I	Média	99,2	105,6	6,4	105,6	6,4	105,6	6,4	99,2	0,0
	Percentil 05	69,5	74,9	7,7	74,9	7,9	75,0	7,9	69,8	0,5
	Percentil 10	75,1	80,7	7,5	80,7	7,5	80,7	7,6	75,3	0,4
	Percentil 25	85,2	91,2	7,1	91,2	7,0	91,2	7,0	85,3	0,2
	Percentil 50	97,7	104,1	6,6	104,0	6,5	104,0	6,5	97,7	0,0
	Percentil 75	111,6	118,3	6,0	118,3	6,1	118,3	6,0	111,5	0,1
	Percentil 90	125,4	132,5	5,7	132,5	5,7	132,5	5,7	125,2	0,2
	Percentil 95	134,3	141,6	5,5	141,6	5,4	141,6	5,5	134,0	0,2
II	Média	98,4	104,7	6,4	104,7	6,5	104,7	6,5	98,4	0,0
	Percentil 05	78,0	85,8	9,9	84,5	8,2	84,5	8,3	78,7	0,8
	Percentil 10	82,1	89,6	9,1	88,5	7,8	88,5	7,9	82,6	0,6
	Percentil 25	89,2	96,2	7,8	95,6	7,2	95,6	7,2	89,5	0,3
	Percentil 50	97,7	104,0	6,5	104,0	6,5	104,0	6,5	97,7	0,0
	Percentil 75	106,8	112,5	5,3	113,1	5,9	113,1	5,9	106,5	0,2
	Percentil 90	115,5	120,6	4,4	121,8	5,4	121,8	5,4	115,1	0,4
	Percentil 95	121,0	125,8	3,9	127,3	5,1	127,3	5,2	120,5	0,5
III	Média	99,2	105,7	6,5	105,7	6,5	105,9	6,7	99,3	0,1
	Percentil 05	69,5	76,5	10,1	75,9	9,2	75,9	9,3	70,4	1,3
	Percentil 10	75,1	82,0	9,3	81,5	8,6	81,5	8,6	75,8	1,0
	Percentil 25	85,2	91,9	7,9	91,6	7,5	91,6	7,6	85,6	0,5
	Percentil 50	97,7	104,4	6,9	104,1	6,6	104,3	6,8	97,7	0,1
	Percentil 75	111,6	117,7	5,5	118,0	5,8	118,2	6,0	111,3	0,2
	Percentil 90	125,4	131,4	4,8	131,9	5,2	132,1	5,4	124,8	0,4
	Percentil 95	134,3	139,6	4,0	141,0	5,0	141,1	5,1	133,6	0,5
IV	Média	98,4	104,8	6,5	104,7	6,5	105,1	6,9	98,5	0,2
	Percentil 05	78,0	87,9	12,6	85,6	9,7	86,1	10,3	78,0	0,0
	Percentil 10	82,1	91,2	11,1	89,3	8,9	89,8	9,4	82,0	0,1
	Percentil 25	89,2	97,0	8,8	96,0	7,6	96,3	8,0	89,1	0,0
	Percentil 50	97,7	104,2	6,7	104,0	6,5	104,3	6,9	97,7	0,1
	Percentil 75	106,8	111,7	4,6	112,7	5,6	113,1	5,9	107,0	0,3
	Percentil 90	115,5	119,2	3,2	121,0	4,8	121,3	5,0	116,1	0,5
	Percentil 95	121,0	123,6	2,1	126,4	4,4	126,5	4,5	121,8	0,7

Abreviações: NCI, *National Cancer Institute*, MSM, *Multiple Source Method*, SPADE, *Statistical Program to Assess Dietary Exposure*, MEID, *Method for estimate intake distribution*.

Tabela 2. EQM das estimativas em relação ao valor verdadeiro para os cenários.

Cenário	Parâmetro	Método			
		MSM	NCI	SPADE	MEID
I	Média	43,5	43,5	43,6	2,9
	Percentil 05	41,5	42,6	42,1	9,8
	Percentil 10	41,6	41,4	41,3	7,4
	Percentil 25	41,1	41,2	40,9	4,1
	Percentil 50	44,1	43,8	43,3	2,8
	Percentil 75	51,2	52,2	51,3	6,4
	Percentil 90	67,3	68,3	66,8	16,7
	Percentil 95	82,6	83,5	81,6	27,7
II	Média	42,2	42,6	42,7	2,4
	Percentil 05	124,3	70,0	71,6	21,0
	Percentil 10	98,3	60,9	61,8	14,4
	Percentil 25	63,5	48,8	49,1	5,9
	Percentil 50	43,1	42,7	42,6	2,3
	Percentil 75	46,5	48,0	47,7	7,7
	Percentil 90	77,6	66,3	66,0	23,5
	Percentil 95	111,2	84,8	84,9	39,3
III	Média	51,1	50,6	108,9	9,0
	Percentil 05	123,8	94,6	94,8	35,7
	Percentil 10	101,0	80,8	80,4	26,5
	Percentil 25	68,4	61,0	60,5	14,1
	Percentil 50	54,7	52,6	105,5	8,8
	Percentil 75	62,2	63,6	111,3	20,7
	Percentil 90	115,9	105,5	147,8	56,0
	Percentil 95	161,2	159,1	190,2	93,6
IV	Média	48,3	48,2	116,7	7,4
	Percentil 05	231,4	139,4	165,2	44,0
	Percentil 10	172,8	109,3	128,8	31,1
	Percentil 25	95,1	69,9	80,7	14,4
	Percentil 50	50,6	48,8	109,1	7,4
	Percentil 75	56,1	58,7	160,6	19,1
	Percentil 90	125,9	105,8	203,4	53,3
	Percentil 95	193,2	160,9	253,3	88,3

Abreviações: NCI, *National Cancer Institute*, MSM, *Multiple Source Method*, SPADE, *Statistical Program to Assess Dietary Exposure*, MEID, *Method for estimate intake distribution*.

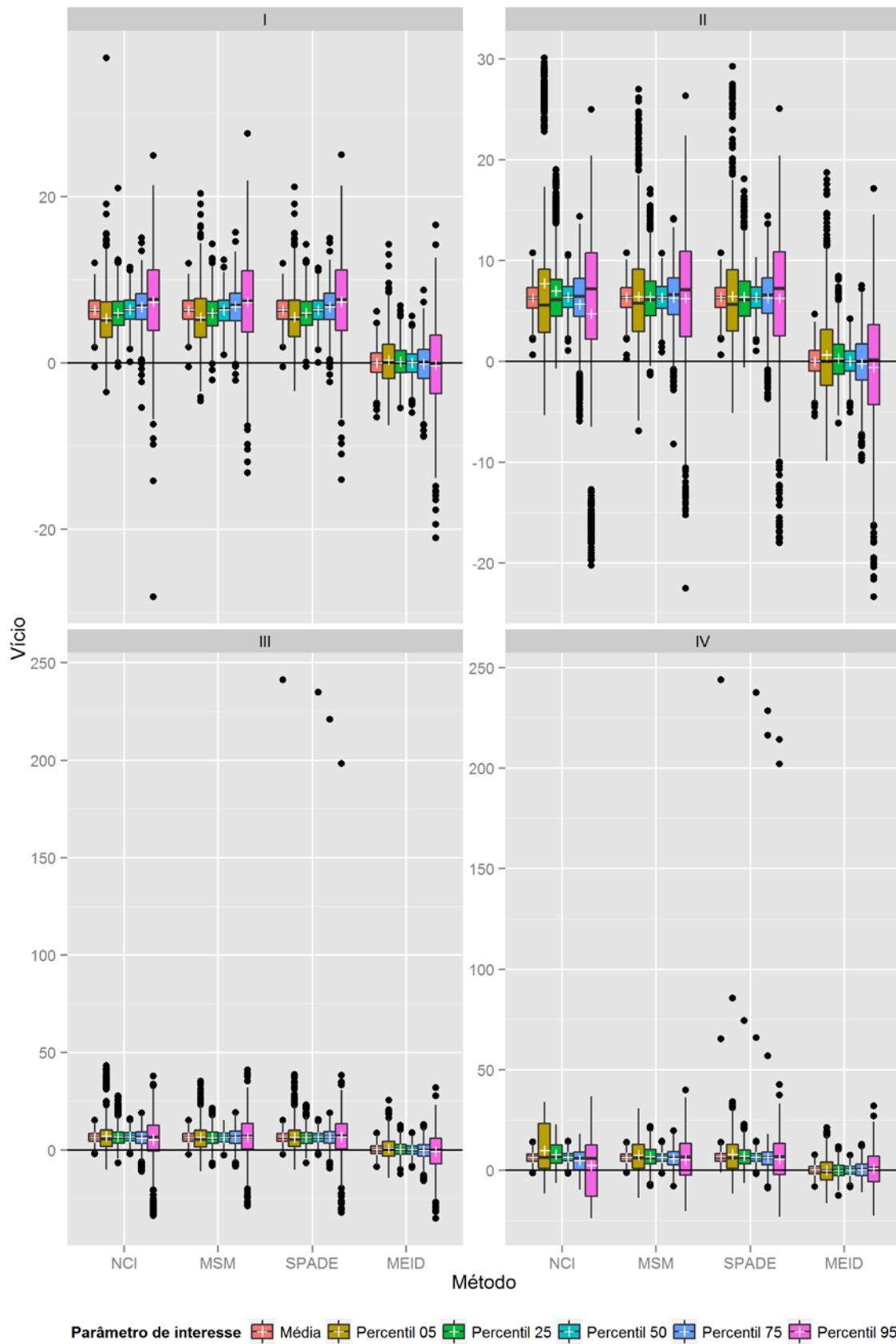


Figura 1. Vícios dos estimadores para todos os cenários.

REFERÊNCIAS

1. Dong J-Y, Zhang L, He K, Qin L-Q. Dairy consumption and risk of breast cancer: a meta-analysis of prospective cohort studies. *Breast Cancer Res Treat.* maio de 2011;127(1):23–31.
2. 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.
3. Gay C. Estimation of population distributions of habitual nutrient intake based on a short-run weighed food diary. *British Journal of Nutrition.* 83º ed 200AD;287–93.
4. 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.
5. 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.
6. 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.
7. 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.
8. 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.
9. 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.
10. 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.
11. 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.
12. 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.
13. 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.
14. 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.
15. Haubrock J, Nöthlings U, Volatier J-L, Dekkers A, Ocké M, Harttig U, et al. Estimating usual food intake distributions by using the multiple source method in the EPIC-Potsdam

- Calibration Study. *J Nutr.* maio de 2011;141(5):914–20.
16. 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, Diplo C, Schwarz N, et al., organizadores. *Survey Measurement and Process Quality* [Internet]. Hoboken, NJ, USA: John Wiley & Sons, Inc.; 2012. p. 689–709. Recuperado de: <http://doi.wiley.com/10.1002/9781118490013.ch30>
 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. The R Project for Statistical Computing [Internet]. Recuperado de: <http://www.r-project.org/>
 19. 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.
 20. 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.
 21. Goedhart PW, Voet H, Sven K, 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 Journal.*
 22. Box GEP, Cox DR. An Analysis of Transformations. *Journal of the Royal Statistical Society Series B (Methodological).* 2º ed 1964;211–52.
 23. Fox J. *An R companion to applied regression.* 2nd ed. Thousand Oaks, Calif: SAGE Publications; 2011. 449 p.
 24. Brown H, Prescott R. *Applied mixed models in medicine.* 2nd ed. Chichester, England □; Hoboken, NJ: John Wiley; 2006. 455 p.
 25. Bates DM, Maechler M, Bolker B. *LME4: Linear mixed-effects models using S4 classes.* setembro de 2013; Recuperado de: <http://cran.r-project.org/web/packages/lme4/>
 26. Stasinopoulos DM, Rigby RA. Generalized Additive Models for Location Scale and Shape GAMLSS in R. *Journal of Statistical Software.* 23º ed dezembro de 2007;
 27. Harttig U, Haubrock J, Knüppel S, Boeing H, EFCOVAL Consortium. 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.
 28. SAS | Business Analytics and Business Intelligence Software [Internet]. Recuperado de: <http://www.sas.com/>
 29. AutoHotkey [Internet]. Recuperado de: <http://www.autohotkey.com/>

Material Suplementar

1) Geração de dados

```
# Pacotes necessários do R
require(lme4)
require(car)
require(gamlss.dist)

#####
# Geração dos dados #####
#####
### Informações iniciais

suj <- 500                                # número de sujeitos
m <- 7.5                                  # média de consumo
varintra <- 1                             # variância intraindividual
rvar <- 4                                  # variância interindividual
varentre <- varintra/rvar                  # variância interindividual
dpintra <- sqrt(varintra)                  # desvio padrão intra
dpentre <- sqrt(varentre)                  # desvio padrão entre
nmax <- suj*2                              # total de observações

# número de repetições (nº de recordatórios)
nrep <- 2
# valor da transformada de Box-Cox
l <- 0.2

### Efeitos aleatórios
set.seed(12356780)                         # semente da simulação
ef_sujeito = rnorm(suj,0,dpentre)           # Efeito aleatório

### Gerando o consumo
media = m + ef_sujeito                       # Gerando média individual com
efeito aleatório
ef_suj_rep = rep(media,each=nrep)           # Efeito do sujeito

### Gerando dois consumos para cada sujeito
x = rnorm(nmax,ef_suj_rep,dpintra)

### Transformando os dados na escala Box-Cox
x2 <- (((1*x)+1)^(1/l))
```

2) Código exemplo do método MEID

```
#####
### Estimação do lambda #
#####

a <- boxCox(x2 ~ 1)                          # Ajusta a Box-Cox

# Escolhe o melhor valor de lambda para transformação dos dados em
# normal
lb <- a$x[which.max(a$y)]

#####
### Transformação dos dados para normal #
#####
```

```

if (lb==0) {dados <- exp(x2)
} else dados <- (((x2^lb)-1)/lb)

#####
### Ajuste do modelo #
#####
# gerando as identificações do sujeito
sujeito = rep(1:suj,each=2)

# criando matriz de consumos transformados e juntando a
identificação # dos sujeitos
dadosyn <- data.frame(dados,sujeito)

# nomeando a matriz
colnames(dadosyn) <- c("consumo","sujeito")

dadosyn$sujeito <- as.factor(dadosyn$sujeito) #criando fator
fit <- lmer(consumo~(1|sujeito),data=dadosyn) # ajustando
modelo

# seleciona a média estimada
inter <- fixef(fit)

# seleciona o desvio padrão entre sujeitos
dp_entre_modelo <- as.numeric(attr(VarCorr(fit)[[1]],'stddev'))

# seleciona o desvio padrão intra sujeitos
dp_intra_modelo <- attr(VarCorr(fit),'sc')

#####
### Cálculo da média e dos percentis #
#####

# desvio padrão do consumo geral transformado da população
dp_box = dp_entre_modelo

# Estimando a média de uma Box-Cox com parâmetros utilizados na
# geração dos dados

integrand1 <- function(x) {x*dLNO(x, nu = lb, mu = inter, sigma =
dp_entre_modelo)}

quantis_box <-
mapply(qLNO,c(0.01,0.05,0.1,0.15,0.25,0.40,0.50,0.75,0.85,0.90,0.95,
0.99), mu = inter, sigma = dp_entre_modelo, nu = lb, lower.tail =
TRUE, log.p = FALSE)

# Estimando a média de uma Box-Cox utilizando os limites de
# integração iguais a Limite superior = 0,00001 e Limite inferior =
# 0,99999

pot <- -5
limite <- qLNO(c(10 ^ pot, 1 - 10 ^ pot), mu = inter, sigma =
dp_entre_modelo, nu = lb, lower.tail = TRUE, log.p = FALSE)
med_box <- integrate(integrand1, lower=limite[1], upper =
limite[2])[1]

```

```
# Reúne as estimativas
Resultados <- c(med_box,quantis_box)

# Nomeia as estimativas
names(Resultados) <- c("Consumo Médio","Percentil 5%","Percentil
10%","Percentil 25%","Percentil 50%","Percentil 75%","Percentil
90%", "Percentil 95%")

# Mostra as Estimativas
(rbind(Resultados))
```

6. CONCLUSÕES E CONSIDERAÇÕES FINAIS

Concluimos que o estimador MEID é acurado e preciso na estimação da distribuição de consumo usual de populações, pois ele obteve os menores vícios e menores EQM quando comparados aos demais. Tem-se também que ele apresentou uma metodologia mais simples que as demais, como também ele pode ser implementado em qualquer *software* que tenha linguagem compatível com a do *software* R.

Apesar dos resultados obtidos é necessário realizar mais estudos de simulação, uma vez que as simulações foram feitas para quatro casos específicos e que haviam sido os mais problemáticos do trabalho de Souverein *et al.* (23). Também, sugere-se o teste dos novos métodos quando se inclui covariáveis, uma vez que a maioria dos métodos permite essa incorporação e não foi avaliada essa característica neste trabalho.

7. REFERÊNCIAS BIBLIOGRÁFICAS

1. Dong J-Y, Zhang L, He K, Qin L-Q. Dairy consumption and risk of breast cancer: a meta-analysis of prospective cohort studies. *Breast Cancer Res Treat.* maio de 2011;127(1):23–31.
2. Carroll RJ, Midthune D, Subar AF, Shumakovich M, Freedman LS, Thompson FE, et al. Taking Advantage of the Strengths of 2 Different Dietary Assessment Instruments to Improve Intake Estimates for Nutritional Epidemiology. *Am J Epidemiol.* 24 de janeiro de 2012;175(4):340–7.
3. Schatzkin A, Kipnis V, Carroll RJ, Midthune D, Subar AF, Bingham S, et al. A comparison of a food frequency questionnaire with a 24-hour recall for use in an epidemiological cohort study: results from the biomarker-based Observing Protein and Energy Nutrition (OPEN) study. *Int J Epidemiol.* 17 de dezembro de 2003;32(6):1054–62.
4. Midthune D, Schatzkin A, Subar AF, Thompson FE, Freedman LS, Carroll RJ, et al. Validating an FFQ for intake of episodically consumed foods: application to the National Institutes of Health-AARP Diet and Health Study. *Public Health Nutr.* julho de 2011;14(7):1212–21.
5. Lissner L, Troiano RP, Midthune D, Heitmann BL, Kipnis V, Subar AF, et al. OPEN about obesity: recovery biomarkers, dietary reporting errors and BMI. *Int J Obes* 2005. junho de 2007;31(6):956–61.
6. Subar AF, Kipnis V, Troiano RP, Midthune D, Schoeller DA, Bingham S, et al. Using intake biomarkers to evaluate the extent of dietary misreporting in a large sample of adults: the OPEN study. *Am J Epidemiol.* 1 de julho de 2003;158(1):1–13.
7. Kipnis V. Structure of Dietary Measurement Error: Results of the OPEN Biomarker Study. *Am J Epidemiol.* 1 de julho de 2003;158(1):14–21.
8. Novotny JA, Rumpler WV, Riddick H, Hebert JR, Rhodes D, Judd JT, et al. Personality characteristics as predictors of underreporting of energy intake on 24-hour dietary recall interviews. *J Am Diet Assoc.* setembro de 2003;103(9):1146–51.
9. Hebert JR, Ebbeling CB, Matthews CE, Hurley TG, MA Y, Druker S, et al. Systematic errors in middle-aged women's estimates of energy intake: comparing three self-report measures to total energy expenditure from doubly labeled water. *Ann Epidemiol.* novembro de 2002;12(8):577–86.
10. Hebert JR, Hurley TG, Peterson KE, Resnicow K, Thompson FE, Yaroch AL, et al. Social desirability trait influences on self-reported dietary measures among diverse participants in a multicenter multiple risk factor trial. *J Nutr.* janeiro de 2008;138(1):226S–234S.

11. 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.
12. 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.
13. 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.
14. 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.
15. 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.
16. 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.
17. 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. Recuperado de: <http://doi.wiley.com/10.1002/9781118490013.ch30>
18. 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 200AD;287–93.
19. 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.
20. 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.
21. 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.
22. Goedhart PW, Voet H, Sven K, Dekkers ALM, Dood KW, Boeing H, et al. A comparison by simulation of different methods to estimate the usual intake distribution for episodically consumed foods. *EFSA Journal.*

23. 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.
24. 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.
25. 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.
26. 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.
27. 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.
28. 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.
29. Box GEP, Cox DR. An Analysis of Transformations. Series B (Methodological). *Journal of the Royal Statistical Society;* 1964;211–52.
30. 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.
31. Pinheiro JC, Bates DM. Approximations to the Log-Likelihood Function in the Nonlinear Mixed-Effects Model. *J Comput Graph Stat.* 1 de março de 1995;4(1):12–35.
32. 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.
33. Ambler G, Benner A. MFP: Multivariable Fractional Polynomials. 2010; Recuperado de: <http://cran.r-project.org/web/packages/mfp/>
34. Scheipl F. AMER: Additive mixed models with lme4. 2011; Recuperado de: <http://cran.r-project.org/web/packages/amer/amer.pdf>

35. Bates DM, Maechler M, Bolker B. LME4: Linear mixed-effects models using S4 classes. setembro de 2013; Recuperado de: <http://cran.r-project.org/web/packages/lme4/>
36. Stasinopoulos DM, Rigby RA. Generalized Additive Models for Location Scale and Shape (GAMLSS) in R. Journal of Statistical Software. 23^o ed dezembro de 2007;