AVALIAÇÃO DE MÉTODOS ESTATÍSTICOS NA ANÁLISE DE DADOS DE CONSUMO ALIMENTAR

Eloisa Elena Paschoalinotte

Dissertação apresentada à Universidade Estadual Paulista "Júlio de Mesquita Filho" para a obtenção do título de Mestre em Biometria.

BOTUCATU
São Paulo - Brasil
Dezembro – 2009

AVALIAÇÃO DE MÉTODOS ESTATÍSTICOS NA ANÁLISE DE DADOS DE CONSUMO ALIMENTAR

Eloisa Elena Paschoalinotte

Orientador: Prof. Dr. **José Eduardo Corrente**

Dissertação apresentada à Universidade Estadual Paulista "Júlio de Mesquita Filho" para a obtenção do título de Mestre em Biometria.

BOTUCATU
São Paulo - Brasil
Dezembro – 2009

Dedicatória

À Deus

Fonte de sabedoria e amor

Nunca teria conseguido sem a sua presença.

À minha querida família,

Meus pais

Maria Inês e Luiz Paschoalinotte

Meu irmão Robrigo e minha avó Joanna.

Agradecimentos

Ao Prof. Dr. José Eduardo Corrente, meu orientador e um amigo, que esteve presente com muito empenho e todo o seu conhecimento durante a realização desse trabalho.

Aos meus pais Luiz e Maria Inês, que sempre me deram todo o apoio necessário, ao carinho do meu irmão Rodrigo e a minha avó Joanna com suas orações.

Aos professores do Departamento de Bioestatística da Universidade Júlio de Mesquista - Unesp - Campos de Botucatu, pela atenção e colaração.

Ao Prof. Dr. Carlos Roberto Padovanni, por ter me dado uma oportunidade de ser sua estagiária no Departamento de Bioestatística e compartilhar de toda a sua experiência nesta área.

Aos funcionários: Bete, Laura, Arthur e Marcos que deram apoio para a realização desse trabalho.

Aos meus amigos do mestrado que me ajudaram durante todo esse processo e especialmente a Giovana pela sua colaboração, incentivo e principalmente por sua amizade.

À minha querida amiga Andreza, que esteve o tempo todo, me incentivando e entendendo os momentos que não pude estar presente. E a todos aqueles amigos e familiares que contribuiram direta ou indiretamente para a realização desse trabalho.

À Deus fonte de esperança, de amor e de vida.

Sumário

Página

LISTA DE FIGURAS **vii**

LISTA DE TABELAS **ix**

RESUMO **xi**

SUMMARY **xiv**

1 INTRODUÇÃO **1**

1.1 Modelo de Regressão Linear e Modelo com Erro de Medida 2

1.2 Relações funcional e estrutural num modelo com erro de medida 5

1.2.1 Relação funcional no modelo linear 5

1.2.2 Relação estrutural no modelo linear 5

1.2.3 Uma Solução dos Mínimos Quadrados 6

1.2.4 Estimação por Máxima Verossimilhança 9

1.3 Métodos Estatísticos para o Consumo Diário de Nutrientes e Alimentos . 11

1.3.1 Modelo com Erro de Medida Clássico 11

1.4 Método do Iowa State University - ISU 15

1.5 Método do Iowa State University Foods - ISUF 16

1.6 Método National Cancer Institute - NCI 18

2 OBJETIVO **21**

3 METODOLOGIA **22**

3.1 Obtenção de dados de consumo nutricional 22
3.2 Aplicação do Metodo ISU 23
3.3 Aplicação do Método NCI 25

4 RESULTADOS E DISCUSSÃO 27

4.1 Análise dos dados do consumo 27
4.2 Resultados da análise dos dados do consumo 36
4.2.1 Análise dos dados do consumo de vitamina B12 para ambos os sexos. . 38
4.3 Resultados das análises dos dados do consumo de nutrientes utilizando o método NCI e incluindo as covariáveis no modelo 41

5 CONCLUSÕES 50

REFERÊNCIAS BIBLIOGRÁFICAS 52

Lista de Figuras

Página

1 A distância minimizada por mínimos quadrados ortogonais. 7

2 Distribuição do consumo de (a) cálcio e (b) zinco para os idosos. Avaré, 2009 30

3 Distribuição do consumo de (c) magnésio e (d) vitamina A para os idosos. Avaré, 2009 30

4 Distribuição do consumo de (e) vitamina C e (f) vitamina E para os idosos. Avaré, 2009 31

5 Distribuição do consumo de vitamina B12 para os idosos. Avaré, 2009 . . 31

6 Distribuição do consumo de (a) cálcio e (b) zinco para os idosos do sexo masculino. Avaré, 2009 32

7 Distribuição do consumo de (c) magnésio e (d) vitamina A para os idosos do sexo masculino. Avaré, 2009 32

8 Distribuição do consumo (e) vitamina C e (f) vitamina E para os idosos do sexo masculino. Avaré, 2009 33

9 Distribuição do consumo de vitamina B12 para os idosos do sexo masculino. Avaré, 2009 33

10 Distribuição do consumo de (a) cálcio e (b) zinco para os idosos do sexo feminino. Avaré, 2009 34

11 Distribuição do consumo de (c) magnésio e (d) vitamina A para os idosos do sexo feminino. Avaré, 2009 34

12 Distribuição do consumo de (e) vitamina C e (f) vitamina E para os idosos do sexo feminino. Avaré, 2009 35

13 Distribuição do consumo de vitamina B12 para os idosos do sexo feminino. Avaré, 2009 35
14 Distribuição do consumo de vitamina B12 para os idosos (a) do sexo masculino e (b) do sexo feminino. Avaré, 2009 38
15 Distribuição do consumo após a transformação raiz quadrada de vitamina B12 para idosos (a) do sexo masculino e (b) do sexo feminino. Avaré, 2009 39

Lista de Tabelas

Página

1 Esquema da ANOVA para o Modelo com Erro de Medida 13

2 Fator de ajuste do viés a ser aplicado na transformação inversa. 15

3 Descrição da amostra de idosos de Avaré, 2009 27

4 Descrição do consumo de nutrientes da amostra de idosos. Avaré, 2009 . 28

5 Descrição do consumo de nutrientes da amostra de idosos estratificado por sexo. Avaré, 2009 29

6 Probabilidade de inadequação de nutrientes para idosos de ambos os sexos pelos métodos ISU e NCI. Avaré, 2009 37

7 ANOVA para o consumo de vitamina B12 pelos idosos do sexo masculino. 39

8 ANOVA para o consumo de vitamina B12 para o sexo feminino. 41

9 Ajuste do modelo com erro de medida proposto pelo método NCI para o consumo de cálcio estratificado por sexo. Avaré, 2009 42

10 Ajuste do modelo com erro de medida proposto pelo método NCI para o consumo de magnésio estratificado por sexo. Avaré, 2009 43

11 Ajuste do modelo com erro de medida proposto pelo método NCI para o consumo de zinco estratificado por sexo. Avaré, 2009. 44

12 Ajuste do modelo com erro de medida proposto pelo método NCI para o consumo de vitamina A estratificado por sexo. Avaré, 2009. 45

13 Ajuste do modelo com erro de medida proposto pelo método NCI para o consumo de vitamina B12 estratificado por sexo. Avaré, 2009. 46

14 Ajuste do modelo com erro de medida proposto pelo método NCI para o consumo de vitamina C estratificado por sexo. Avaré, 2009. 47

15 Ajuste do modelo com erro de medida proposto pelo método NCI para o consumo de vitamina E estratificado por sexo. Avaré, 2009. 49

AVALIAÇÃO DE MÉTODOS ESTATÍSTICOS NA ANÁLISE DE DADOS DE CONSUMO ALIMENTAR

Autora: ELOISA ELENA PASCHOALINOTTE
Orientador: Prof. Dr. JOSÉ EDUARDO CORRENTE

RESUMO

A avaliação do consumo alimentar de um indivíduo ou de uma população tem sido um desafio tanto para profissionais da área de nutrição como para os da área de estatística. Isso porque a característica central do consumo alimentar é a variabilidade da dieta, que pode gerar uma grande variabilidade inter e intrapessoal. Para contornar esse problema, métodos estatísticos apropriados foram desenvolvidos baseados no modelo de regressão com erro de medida de modo a se obter a distribuição estimada do consumo habitual. Dentre os métodos de avaliação de consumo, têm-se o método da *Iowa State University* (ISU), o método da *Iowa State University for Foods* (ISUF) e o método do *National Cancer Institute* (NCI). Todos esses métodos são baseados no modelo com erro de medida incorporando a questão do consumo esporádico (método ISUF) e a possibilidade de incluir covariáveis que podem interferir na distribuição estimada do consumo (método NCI). Para o uso do

método ISU, foi desenvolvido um programa chamado PC-SIDE (*Software for Intake Distribution Estimate*), que fornece a distribuição do consumo habitual bem como a probabilidade de inadequação de determinados nutrientes de acordo com as recomendações nutricionais. Com esse mesmo programa, é possível obter a distribuição do consumo habitual esporádico, dado pelo método ISUF. Para o método NCI, foram desenvolvidas macros no programa SAS (*Statistical Analysis System*) que permitem incluir covariáveis e estimar a distribuição do consumo habitual baseado no modelo com erros de medidas. Desse modo, o objetivo deste trabalho foi avaliar essas metodologias estatísticas na análise de dados de consumo alimentar e aplicá-los a um conjunto de dados de um levantamento nutricional de idosos. Foram estudadas as metodologias de ajuste dos modelos propostos para a obtenção da distribuição estimada de consumo baseado nos métodos ISU, ISUF e NCI. Os métodos ISU e NCI foram aplicados os dados de três recordatórios 24-horas obtidos de uma pesquisa sobre a qualidade de vida de idosos na cidade de Avaré, SP no ano de 2008. A amostra inicial foi composta de 365 idosos, e uma subamostra de 20% (73 idosos) foi obtida para realizar o levantamento nutricional. Foi considerado o consumo de cálcio, zinco, magnésio, vitamina A, vitamina B12, vitamina C e vitamina E, e como covariáveis, consideram-se os aspectos sociodemográficos, de prevenção e morbidades. A distribuição do consumo habitual foi obtida levando-se em conta a subamostra total e estratificada por sexo. A probabilidade de inadequação no consumo desses nutrientes pelos idosos também foi obtida pelos dois métodos, utilizando-se as recomendações de consumo para esses nutrientes. Verificou-se que, para os nutrientes considerados, a probabilidade de inadequação de consumo foi alta para este grupo de idosos, sendo, em alguns casos, muito diferente quando estimada pelos métodos ISU e NCI, mesmo quando as covariáveis foram não significativas. Conclui-se que mais estudos são necessários, talvez considerando amostras maiores ou outras covariáveis de maior interesse, para que se possam obter melhores estimativas para distribuição do consumo habitual e, consequentemente, para estimar a probabilidade de inadequação no consumo de nutrientes.

Palavras-chave: modelo com erro de medida, probabilidade de inadequação, consumo de idosos, ajuste para dados de consumo

EVALUATION OF STATISTICAL METHODS IN FOOR INTAKE DATA ANALYSIS

Author: ELOISA ELENA PASCHOALINOTTE
Adviser: Prof. Dr. JOSÉ EDUARDO CORRENTE

SUMMARY

Evaluating an individual's or a population's food intake has been a challenge for both nutrition professionals and statisticians. This is so because the main characteristic of food intake is diet variety, which may generate large between- and within-person variability. In order to overcome this problem, appropriate statistical methods have been developed based on the measurement-error regression model so as to obtain the estimated distribution of usual intake. Among the intake evaluation methods are the Iowa State University (ISU), the Iowa State University for Foods (ISUF) and the National Cancer Institute (NCI) methods. All of them are based on the measurement- error model incorporating the issue concerning episodic intake (ISUF method) and the possibility of including covariates that can interfere in the intake estimated distribution (NCI method). In order to use the ISU method, a software referred to as PC-SIDE (Software for Intake Distribution Estimate) was

designed. It provides the usual intake distribution as well as the probability of inadequacy for certain nutrients according to nutritional recommendations. By using the same software, it is possible to obtain the distribution of episodic usual intake given by the ISUF method. For the NCI method, macros were developed in the SAS (Statistical Analysis System) software which enable the inclusion of covariates and the estimation of the usual intake distribution based on the measurement-error model. Hence, this study aimed at evaluating these statistical methodologies in the analysis of food intake data and at applying them to a data set for a nutritional assessment of elderly individuals. The fitting methodologies for the models proposed to obtain the estimated intake distribution based on the ISU, ISUF and NCI methods were studied. The ISU and NCI methods were applied to data from three 24-hours recalls obtained from a study on the quality of life of elderly persons in the city of Avaré - SP in 2008. The initial sample consisted of 365 elderly individuals, and a sub-sample of 20% (73 elderly persons) was obtained to conduct the nutritional assessment. The intake of calcium, zinc, magnesium, vitamin A, vitamin B12, vitamin C and vitamin E was considered as well as the individuals' sociodemographic aspects, prevention aspects and morbidities as covariates. The usual intake distribution was obtained by taking into account the total sub-sample and the sub-sample stratified by gender. The probability of inadequacy in the intake of these nutrients by the elderly persons was also obtained by the two methods using their intake recommendations. It was observed that, for the nutrients under consideration, the probability of intake inadequacy was high in this group of individuals, and, in some cases, it was very different when estimated by the ISU and NCI methods, even when the covariates were non-significant. It is concluded that further studies are required, maybe considering larger samples or other covariates of greater interest, in order to obtain better estimates for usual intake distribution and, consequently, to estimate the probability of inadequacy in nutrient intake.

Key words: measurement-error model, probability of inadequacy, elderly individuals' intake, intake data fitting.

1 INTRODUÇÃO

A avaliação do consumo alimentar tem um papel fundamental na área de pesquisa e saúde para o desenvolvimento de recomendações para uma alimentação saudável de indivíduos e grupos. De acordo com Willett (1998), a característica central do consumo alimentar de um indivíduo ou de uma população sadia é a variabilidade diária da dieta. Entretanto, são necessários métodos apropriados para avaliar a ingestão alimentar de indivíduos e/ou população.

A maneira mais frequentemente usada na obtenção de dados de consumo alimentar é o Recordatório 24 horas (R24h), que consiste em definir e quantificar todos alimentos e bebidas ingeridos no período anterior à entrevista, que pode ser as 24 horas precedentes ou, mais comumente, o dia anterior Gibson (1990). De acordo com Fisberg et al. (2005), trata-se de uma entrevista pessoal em profundidade e conduzida por um entrevistador treinado, podendo também ser realizada por telefone. É possível ainda que o indivíduo responda a seu próprio recordatório, listando os alimentos e as bebidas que consumiu. Os pesquisadores da área de nutrição reconhecem a necessidade de múltiplos recordatórios para estimar o consumo habitual, que é definido como a média do consumo alimentar em um período de tempo determinado (meses ou um ano), em que o indivíduo mantém um padrão constante. Nota-se, então, que o consumo habitual não seria possível obter com apenas um recordatório, que forneceria uma estimativa pouco confiável para o mesmo. Mas, mesmo com um número considerável de recordatórios, as médias de consumo ainda contém considerável variação intrapessoal e, estimar a proporção da população com consumo habitual acima ou abaixo de um valor de referência, pode conduzir a um estimador inadequado já que a avaliação do consumo adequado freqüentemente en-

volve padrões que pertencem às caudas da distribuição, contendo vieses substanciais (Dodd & Guenther, 2006).

A variação nos consumos diários é composta por dois componentes. Um componente é a dificuldade de quantificar corretamente as quantidades consumidas atualmente. Outro componente da variação do consumo diário está associado com a variabilidade diária do consumo alimentar individual. A soma destes dois componentes produz uma grande variância intrapessoal (Hegsted, 1972b,a; Beaton et al., 1972; Council, 1986; Nusser et al., 1996a). Com isto, pode-se observar a relação existente entre a variação sistemática associada ao dia da semana, ou dia da entrevista, dentre outros fatores. Por isso, a avaliação de consumo tem sido um desafio para nutricionistas e estatísticos, já que o uso de metodologias de análise devem ser apropriadas para se obter uma estimativa adequada do consumo habitual. E, uma das maneiras de avaliação do consumo é o uso de modelos de regressão com erros de medidas como uma forma de incluir efeitos que possam fornecer uma melhor estimativa do consumo habitual.

A seguir, será descrito o modelo de regressão linear e o modelo com erros de medidas , e os métodos usuais de consumo.

1.1 Modelo de Regressão Linear e Modelo com Erro de Medida

Sejam n pares de duas variáveis, (X_i, Y_i) para $i = 1, 2, \ldots, n$. De acordo com Casella & Berger (2001), num modelo de regressão linear simples, tem-se uma relação da forma:

$$Y_i = \alpha + \beta X_i + \epsilon_i, \qquad \text{para } i = 1, \ldots, n, \tag{1}$$

em que Y_i é uma variável aleatória e X_i é uma variável observável ou também chamada de explanatória. As quantidades α e β são o intercepto e a inclinação, considerados como parâmetros fixos e desconhecidos do modelo de regressão.

Para um dado valor de X_i, a variância do erro ϵ_i é necessariamente uma variável aleatória com $E(\epsilon_i) = 0$ e $var(\epsilon_i) = \sigma^2$. Outra suposição comumente feita é que $\epsilon_i \sim \text{Normal}(0, \sigma^2)$, para $i = 1, \ldots, n$. Com isto, de (1) tem-se que

$$E(Y_i) = \alpha + \beta X_i, \qquad \text{para } i = 1, \ldots, n,$$

$$\text{e} \quad Y_i \sim N(\alpha + \beta X_i, \sigma^2).$$

Na prática, considerando n pares de observações $(x_1, y_1), \ldots, (x_n, y_n)$, o interesse será saber se a relação entre x_i e y_i, é linear, ou seja:

$$y_i = \alpha + \beta x_i + \epsilon_i, \qquad \text{para } i = 1, \ldots, n, \tag{2}$$

em que y_i é o valor da variável resposta, α e β são os parâmetros desconhecidos, x_i é uma variável explanatória e ϵ_i são os erros de ajuste, que se supõe seguirem uma distribuição normal com média zero e variância σ^2. A estimação dos parâmetros α e β pode ser feita pelo método de mínimos quadrados, que consiste em minimizar a soma de quadrados dos erros. Este método de estimação leva em conta apenas os parâmetros de interesse, não sendo possível a estimação de σ^2. Com a suposição de que os erros seguem uma distribuição normal, pode-se estimar σ^2 pelo método da máxima verossimilhança.

Porém, ainda de acordo Casella & Berger (2001), existem algumas extensões deste modelo que são úteis em problemas práticos, no qual pode-se considerar modelos com erros no preditor, que é conhecida como regressão com erros na variável (EIV - *errors in variables*). Nestes modelos, a variável X se torna aleatória tanto quanto Y.

A Regressão com Erros na Variável ou nas Variáveis (EIV) é também conhecida como Modelo com Erro de Medida e é fundamentalmente diferente da regresão linear simples, sendo uma generalização do modelo de regressão usual. Entretanto, os problemas que normalmente aparecem com este tipo de modelo são muito diferentes.

Voltando ao modelo de regressão linear simples, considere os pares de variáveis aleatórias (X_i, Y_i), e o modelo da forma:

$$Y_i = \alpha + \beta X_i + \epsilon_i, \quad \text{para } i = 1, \ldots, n,$$

em que X_i seja aleatória com média x_i, para $i = 1, 2, \ldots, n$. Com isto, tem-se que a variável aleatória X_i não é mais medida com média x_i, mas sim, com média ξ_i.

Tem-se, então, que no modelo EIV geral, são observados os pares (x_i, y_i) amostrados de variáveis aleatórias (X_i, Y_i) cuja médias satisfazem a relação linear:

$$E(Y_i) = \alpha + \beta(E(X_i)). \tag{3}$$

Definindo:

$$E(Y_i) = \eta_i \quad \text{e} \quad E(X_i) = \xi_i,$$

tem-se que:

$$\eta_i = \alpha + \beta \xi_i,$$

que é uma relação linear entre as médias da variáveis aleatórias X_i e Y_i, para $i = 1, 2, \ldots, n$. As variáveis ξ_i e η_i são também chamadas de variáveis latentes.

O modelo especificado em (3), não faz distinção entre X e Y. Cabe, então, ao pesquisador especificar Y como sendo a variável resposta, X a variável preditora, e definir o modelo com erro de medida. Desse modo, observa-se os pares (X_i, Y_i), $i = 1, 2, \ldots, n$, independentemente de acordo com

$$Y_i = \alpha + \beta \xi_i + \epsilon_i, \qquad \epsilon_i \sim N(0, \sigma_\epsilon^2), \quad \text{para} \quad i = 1, 2, \ldots, n, \tag{4}$$

$$X_i = \xi_i + \delta_i, \qquad \delta_i \sim N(0, \sigma_\delta^2), \quad \text{para} \quad i = 1, 2, \ldots, n. \tag{5}$$

1.2 Relações funcional e estrutural num modelo com erro de medida

1.2.1 Relação funcional no modelo linear

De acordo com Casella & Berger (2001), existem duas relações diferentes que podem ser especificadas num modelo EIV: uma que especifica a relação linear *funcional* e outra que descreve a relação *estrutural* do modelo. As diferentes especificações podem conduzir a diferentes estimadores com diferentes propriedades.

No modelo apresentado em (4) e (5), no qual se tem as variáveis aleatórias X_i e Y_i com $E(Y_i) = \eta_i$ e $E(X_i) = \xi_i$, supõe-se que a relação funcional seja

$$\eta_i = \alpha + \beta\xi_i, \quad \text{para} \quad i = 1, 2, \ldots, n.$$

Observa-se, agora os pares (X_i, Y_i), $i = 1, 2, \ldots n$, de acordo com

$$Y_i = \alpha + \beta\xi_i + \epsilon_i, \qquad \epsilon_i \sim N(0, \sigma_\epsilon^2),$$

$$X_i = \xi_i + \delta_i, \qquad \delta_i \sim N(0, \sigma_\delta^2),$$

em que ξ_i são parâmetros fixados e desconhecidos e ϵ_i e δ_i são independentes, para $i = 1, 2, \ldots, n$. Os parâmetros de maior interesse são α e β, e a inferência para estes parâmetros é feita usando a distribuição conjunta de (X_i, Y_i), condicionada a ξ_i, para $i = 1, 2, \ldots, n$.

1.2.2 Relação estrutural no modelo linear

Este modelo pode ser visto como uma extensão da relação funcional do modelo. Da mesma forma que na relação funcional, tem-se as variáveis aleatórias X_i e Y_i com $E(X_i) = \xi_i$, $E(Y_i) = \eta_i$, supondo a relação funcional $\eta_i = \alpha + \beta\xi_i$, para $i = 1, 2, \ldots, n$. A diferença é que, agora, os parâmetros ξ_i são variáveis aleatórias, para $i = 1, 2, \ldots, n$. Daí, condicionado a ξ_i, observa-se os pares (X_i, Y_i), $i = 1, 2, \ldots, n$, de acordo com:

$$Y_i = \alpha + \beta\xi_i + \epsilon_i, \qquad \epsilon_i \sim N(0, \sigma_\epsilon^2),$$

$$X_i = \xi_i + \delta_i, \qquad \delta_i \sim N(0, \sigma_\delta^2),$$

e também

$$\xi_i \sim N(0, \sigma_\xi^2).$$

A inferência, neste casos, é feita usando a distribuição conjunta de $((X_i, Y_i), ..., (X_n, Y_n))$ não condicionada aos $\xi_1, ..., \xi_n$.

Os modelos funcional e estrutural, são praticamente similares nas propriedades dos estimadores, ou seja, estimadores que são consistentes no modelo funcional são também consistentes no modelo estrutural. Isso acontece devido ao modelo funcional ser uma versão condicional do modelo estrutural.

1.2.3 Uma Solução dos Mínimos Quadrados

Afim de encontrar a melhor reta através dos pontos observados (x_i, y_i) para $i = 1, 2, \ldots, n$, tem-se que, se x_i for considerado sem erro, faz sentido considerar as menores distâncias verticais do valor observado em relação à reta ajustada. Esta distância supõe que o valor x_i é correto e resulta em Mínimos Quadrados Ordinários. Mas o fato é que, existem alguns problemas no Método de Minímos Quadrados Ordinários com relação aos modelos com erro de medida. Não há razão para se considerar as distâncias verticais, quando os x_i, são considerados com erros associados. Uma solução neste caso, é utilizar o Método de Mínimos Quadrados Ortogonais, que consiste em minimizar as distâncias ortogonais entre o ponto observado e a reta a ser ajustada. Esta distância não irá a favor da variável x_i, como em mínimos quadrados ordinários, mas considera-se ambas as variáveis. A figura 1 mostra a diferença entre os mínimos quadrados ordinários e os mínimos quadrados ortogonais para minimizar a distância de um ponto $P(x^{'}, y^{'})$ a uma reta $y = a + bx$.

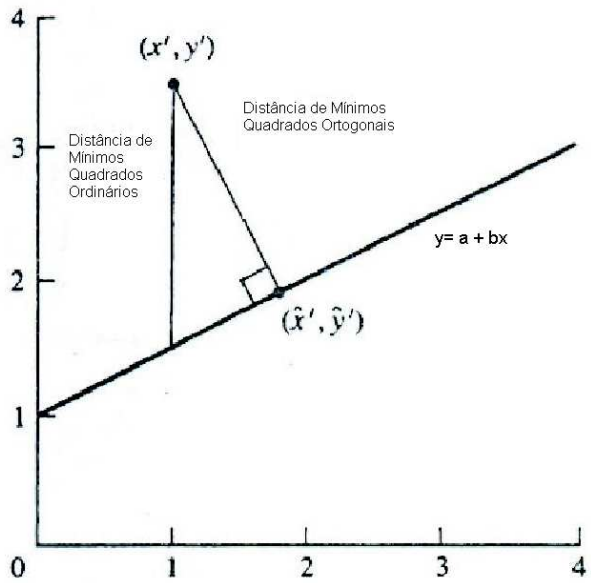


Figura 1 - A distância minimizada por mínimos quadrados ortogonais.

Para calcular a distância de um ponto $P(x^{'}, y^{'})$ à reta dada, utiliza-se algumas ferramentas matemáticas. Inicialmente, tem-se que a distância do ponto $P(x^{'}, y^{'})$ ao ponto $P_i(\widehat{x}^{'}, \widehat{y}^{'})$ é dada por:

$$d(P, P_i) = \sqrt{(x' - \widehat{x}')^2 + (y' - \widehat{y}')^2}. \tag{6}$$

Traçar uma reta passando pelo ponto $P(x^{'}, y^{'})$, paralelamente ao eixo x. Desse modo, constroi-se um triângulo retângulo e pode-se utilizar a teoria de triângulos semelhantes para calcular a distância ortogonal. Assim, tem-se:

$$d(P, y) = \frac{|y^{'} - a - bx^{'}|}{\sqrt{1 + b^2}}, \tag{7}$$

em que $y = a + bx$. Tem-se que (6) e (7) representam a mesma distância. Igualando os valores, tem-se:

$$\sqrt{(x' - \widehat{x}')^2 + (y' - \widehat{y}')^2} = \frac{|y^{'} - a - bx^{'}|}{\sqrt{1 + b^2}}. \tag{8}$$

Resolvendo a igualdade em função de $\widehat{x}^{'}$ e $\widehat{y}^{'}$, tem-se

$$\widehat{x}^{'} = \frac{by^{'} + x^{'} - ab}{1 + b^2} \qquad \text{e} \qquad \widehat{y}^{'} = a + \frac{b}{1 + b^2}(by^{'} + x^{'} - ab). \tag{9}$$

Considerando, agora, os valores observados (x_i, y_i), a distância quadrática entre um ponto observado (x_i, y_i) e um ponto na reta $\widehat{y} = a + b\widehat{x}$ é $(x_i - \widehat{x_i})^2 + (y_i - \widehat{y_i})^2$, em que $\widehat{x}_i$ e $\widehat{y}_i$ são definidos pelas equações (9), para $i = 1, 2, \ldots, n$. O problema consiste, então, em obter a e b, minimizando:

$$S = \sum_{i=1}^{n}((x_i - \widehat{x_i})^2 + (y_i - \widehat{y_i})^2). \tag{10}$$

Considerando que $\widehat{x_i} = \widehat{x}^{'}$, e $\widehat{y_i} = \widehat{y}^{'}$, definidos nas equações (9), e substituindo os valores em (10) tem-se:

$$S = \sum_{i=1}^{n}(\frac{b^2}{(1 + b^2)^2}[y_i - (a + bx_i)]^2 + \frac{1}{(1 + b^2)^2}[y_i - (a + bx_i)]^2) =$$

$$= \frac{1}{1 + b^2}\sum_{i=1}^{n}(y_i - (a + bx_i))^2. \tag{11}$$

Portanto, minimizando em função de a, tem-se, $a = \overline{y} - b\overline{x}$. E substituindo $a = \overline{y} - b\overline{x}$ na equação (11), a solução dos mínimos quadrados totais é a solução que minimiza b, assim:

$$S = \frac{1}{1 + b^2}\sum_{i=1}^{n}((y_i - \overline{y}) - b(x_i - \overline{x}))^2. \tag{12}$$

Desse modo, as somas dos quadrados e dos produtos cruzados são dados por:

$$S_{xx} = \sum_{i=1}^{n}(x_i - \overline{x})^2, \qquad S_{yy} = \sum_{i=1}^{n}(y_i - \overline{y})^2, \qquad S_{xy} = \sum_{i=1}^{n}(x_i - \overline{x})(y_i - \overline{y}).$$

Expandindo o quadrado da equação (12) e substituindo as somas de quadrados e dos produtos cruzados, tem-se:

$$S = \frac{1}{1 + b^2}[S_{yy} - 2bS_{xy} + b^2S_{xx}].$$

Portanto, encontra-se a reta dos mínimos quadrados ortogonais dado pelo reta $y = a + bx$, com

$$a = \overline{y} - b\overline{x} \quad \text{e} \quad b = \frac{-(S_{xx} - S_{yy}) + \sqrt{(S_{xx} - S_{yy})^2 + 4(S_{xy})^2}}{2S_{xy}}. \tag{13}$$

Como já era esperado, esta reta é diferente da reta de mínimos quadrados ordinários. Neste caso, obtivemos uma solução de mínimos quadrados matemática para ajustar a reta. Porém, quando buscamos uma solução estatística, algumas dificuldades são encontradas na estimação de verossimilhança.

1.2.4 Estimação por Máxima Verossimilhança

Consideremos, agora, a obtenção dos estimadores de α e β pelo método de máxima verossimilhança para o modelo linear de relação funcional, pois, a situação para o modelo linear de relação estrutural, é similar. Supondo que X_i e Y_i seguem uma distribuição normal, tem-se que a relação funcional pode ser expressa por:

$$Y_i \sim N(\alpha + \beta\xi_i, \sigma_\epsilon^2) \quad \text{e} \quad X_i \sim N(\xi_i, \sigma_\delta^2),$$

em que os X_i e Y_i são independentes, para $i = 1, 2, \ldots n$. Desse forma, dadas as observações $(\mathbf{x}, \mathbf{y}) = ((x_1, y_1), \ldots, (x_n, y_n))$, tem-se que a função de verossimilhança é dada por:

$$L(\alpha, \beta, \xi_i, ..., \xi_n, \sigma_\delta^2, \sigma_\epsilon^2 | (x, y)) =$$

$$= \frac{1}{(2\pi)^n} \frac{1}{(\sigma_\delta^2 \sigma_\epsilon^2)^{\frac{n}{2}}} \exp\left[-\sum_{i=1}^{n} \frac{(x_i - \xi_i)^2}{2\sigma_\delta^2}\right] \exp\left[-\sum_{i=1}^{n} \frac{(y_i - (\alpha + \beta\xi_i))^2}{2\sigma_\epsilon^2}\right].$$

O problema com esta função de verossimilhança é que ela não possui um máximo finito. Uma maneira de compreender este fato, seria fazer $\xi_i = x_i$ e $\sigma_\delta^2 \to 0$. O valor da função tende ao infinito, mostrando que não há solução de máxima verossimilhança. Solari (1969) mostrou que as equações obtidas da primeira derivada de L e igualando a zero, resultam num ponto de cela e não em um máximo. Quando há um controle total sobre os parâmetros, pode-se forçar a função de verossimilhança

tender ao infinito. Em particular, pode-se fazer a variância tender a zero, enquanto se mantem o termo exponencial limitado.

Casella & Berger (2001) considera uma suposição comum, que não é somente razoável, mas auxilia em muitos problemas, como, por exemplo, fazer $\sigma_\delta^2 = \lambda\sigma_\epsilon^2$ para $\lambda > 0$, fixado e conhecido. Sob esta suposição, pode-se escrever a função de verossimilhança da seguinte forma:

$$L(\alpha, \beta, \xi_i, ..., \xi_n, \sigma_\delta^2|(x, y)) =$$

$$= \frac{1}{(2\pi)^n} \frac{\lambda^{\frac{n}{2}}}{(\sigma_\delta^2)^n} \exp\left[-\sum_{i=1}^{n} \frac{(x_i - \xi_i)^2 + \lambda(y_i - (\alpha + \beta\xi_i))^2}{2\sigma_\delta^2}\right],$$

que agora pode ser maximizada. A maximização é feita em dois estágios. No primeiro estágio, para cada valor de α, β e σ_δ^2, maximizar L com relação a $\xi_1, \ldots, \xi_n$, é o mesmo que maximizar $\sum_{i=1}^{n}((x_i - \xi_i)^2 + \lambda(y_i - (\alpha + \beta\xi_i))^2)$. Para cada $i = 1, 2, \ldots, n$, o mínimo é dado por

$$\xi_i^* = \frac{x_i + \lambda\beta(y_i - \alpha)}{1 + \lambda\beta^2}. \tag{14}$$

Substituindo n valores de ξ_i^*, tem-se

$$\sum_{i=1}^{n}((x_i - \xi_i^*)^2 + \lambda(y_i - (\alpha + \beta\xi_i^*))^2) = \frac{\lambda}{1 + \lambda\beta^2} \sum_{i=1}^{n}(y_i - (\alpha + \beta x_i))^2.$$

Desse modo, a função de verossimilhança agora é dada por:

$$max_{\xi_1,\ldots,\xi_n} L(\alpha, \beta, \xi_i, \ldots, \xi_n, \sigma_\delta^2|(x, y)) =$$

$$= \frac{1}{(2\pi)^n} \frac{\lambda^{\frac{n}{2}}}{(\sigma_\delta^2)^n} \exp\left\{-\frac{1}{2\sigma_\delta^2}\left[\frac{\lambda}{1 + \lambda\beta^2} \sum_{i=1}^{n}(y_i - (\alpha + \beta x_i))^2\right]\right\}. \tag{15}$$

Para maximizar (15), define-se:

$$\alpha^* = \sqrt{\lambda\alpha}, \qquad \beta^* = \sqrt{\lambda\beta}, \qquad y_i^* = \sqrt{\lambda y_i}, \quad \text{para} \qquad i = 1, \ldots, n, \tag{16}$$

o que torna o expoente da forma:

$$\frac{\lambda}{1 + \lambda\beta^2} \sum_{i=1}^{n}(y_i - (\alpha + \beta x_i))^2 = \frac{1}{1 + \beta^{*2}} \sum_{i=1}^{n}(y_i^* - (\alpha^* + \beta^* x_i))^2, \tag{17}$$

que, é idêntica à expressão dos mínimos quadrados ortogonais. De (13) encontram-se os valores minimizados de α^* e β^*, e usando (16) obtém-se a estimação por máxima verossimilhança para α e β, ou seja,

$$\widehat{\alpha} = \overline{y} - \beta\overline{x} \quad \text{e} \quad \widehat{\beta} = \frac{-(S_{xx} - \lambda S_{yy}) + \sqrt{(S_{xx} - \lambda S_{yy})^2 + 4\lambda(S_{xy})^2}}{2\lambda S_{xy}}. \tag{18}$$

Quando $\lambda = 1$, o estimador de máxima verossimilhança coincide com o estimador de mínimos quadrados ortogonais. Combinando as equações (18) com (15), tem-se:

$$max_{\xi_1,...,\xi_n} L(\alpha, \beta, \xi_i, ..., \xi_n, \sigma_\delta^2 | (x, y)) =$$
$$= \frac{1}{(2\pi)^n} \frac{\lambda^{\frac{n}{2}}}{(\sigma_\delta^2)^n} \exp\left[-\frac{1}{2\sigma_\delta^2} \frac{\lambda}{1 + \lambda\beta^2} \sum_{i=1}^{n} (y_i - (\widehat{\alpha} + \widehat{\beta} x_i))^2\right].$$

Para completar a maximazação de L, basta maximizá-la para σ_δ^2, cuja solução é dada por:

$$\widehat{\sigma_\delta^2} = \frac{1}{2n} \frac{\lambda}{1 + \lambda\widehat{\beta^2}} \sum_{i=1}^{n} (y_i - (\widehat{\alpha} + \widehat{\beta} x_i))^2.$$

Pela propriedade de invariância a estimação de máxima verossimilhança, segue que $\widehat{\sigma}_\epsilon^2 = \frac{\widehat{\sigma_\delta^2}}{\lambda}$ e $\widehat{\xi}_i = \widehat{\alpha} + \widehat{\beta} x_i$. É importante ressaltar que $\widehat{\alpha}$ e $\widehat{\beta}$ são estimadores consistentes, mas $\widehat{\sigma}_\delta^2$ não.

Uma forma análoga pode ser obtido para o modelo linear com relação estrutural. Os detalhes podem ser vistos em Casella & Berger (2001).

1.3 Métodos Estatísticos para o Consumo Diário de Nutrientes e Alimentos

1.3.1 Modelo com Erro de Medida Clássico

De acordo com Tooze et al. (2006b) a proposta de um levantamento dietético é avaliar o consumo de um grupo ou de uma população relativo a algum padrão referente a adequação de nutrientes e à prevenção de doenças crônicas. Os

instrumentos mais utilizados para a coleta de dados nutricionais são o Recordatório 24 horas (R24h) e o Questionário de Frequência Alimentar (QFA). O R24h fornece muitos detalhes sobre os tipos e quantidades de alimentos consumidos. Contudo, existe uma grande variabilidade na dieta dos indivíduos no dia-a-dia acumulando erros com o uso de receitas padronizadas e as tabelas de composição dos alimentos. Isto faz com que exista uma considerável variabilidade intrapessoal, caracterizando que um único R24h não fornece uma boa estimativa do consumo habitual. Os modelos estastísticos para dados de consumo alimentar visam atenuar os efeitos das limitações dos R24h, removendo o efeito de variabilidade intrapessoal.

A seguir, são apresentados os métodos utilizados para a análise de dados nutricionais, em que cada método segue a seguinte sequência de passos.

Passo 1: Descreve a relação entre as medidas do R24h e o consumo habitual do indivíduo;

Esta suposição é equivalente a dizer que um R24h é não viesado para o consumo habitual de um único dia. Porém, a falta de viés não implicará na falta de erro. Isto significa que um particular R24h pode super ou subestimar o verdadeiro consumo habitual, mas, em aplicações repetidas para o mesmo indivíduo, os erros estimados podem ser cancelados.

Passo 2: A partição da variação total dos R24h em duas componentes de variância; a variância intra e interpessoal.

O consumo habitual pode ser expresso como a soma da média do consumo habitual do grupo e os desvios individuais em relação ao grupo. Estes desvios representam a variação interpessoal. Logo, o consumo habitual pode ser expresso como:

[I] Consumo habitual= média do consumo habitual do grupo + (consumo habitual - média do consumo habitual do grupo).

Assim, cada recordatório pode ser expresso como:

[II] Consumo R24h= média do consumo habitual do grupo + (consumo habitual - média do consumo habitual do grupo)+ (consumo R24h - consumo

habitual).

É importante observar o terceiro termo da segunda equação, representa a variação intrapessoal. Para estimá-la, é necessário repetidas medidas do R24h, ou seja, pelo menos dois recordatórios.

Para a obtenção dos componentes de variância inter e intrapessoal, considere R_{ij} e T_{ij} como sendo o consumo verdadeiro de um nutriente ou alimento e o correspondente R24h, para o indivíduo i, $i = 1, 2, \ldots, n$, no dia j, $j = 1, 2, \ldots, j_i$, isto é:

$$R_{ij} = T_i + \epsilon_{ij},$$

em que ϵ_{ij} representam os erros intrapessoais que contém a variação diária do consumo individual e outras fontes de erros associados.

Suponha que $T_i \sim \text{Normal}(\mu_w, \sigma_w^2)$ e $\epsilon_{ij} \sim \text{Normal}(0, \sigma_\epsilon^2)$, $i = 1, 2, \ldots, n$, $j = 1, 2, \ldots, j_i$. Supondo $k = max\{j_i\}$, $i = 1, 2, \ldots, n$, a Análise de Variância associada ao modelo proposto é dada pela Tabela (1)

Tabela 1. Esquema da ANOVA para o Modelo com Erro de Medida

F. Variação	GL	SQ	$E(QM)$
Interpessoal	$n-1$	$k\sum_{i=1}^{n}(\overline{R}_{j.} - \overline{R}_{..})^2$	$\sigma_w^2 + k\sigma_b^2$
Intrapessoal	$n(k-1)$	$\sum_{i=1}^{n}\sum_{j=1}^{k}(R_{ij} - \overline{R}_{i.})^2$	σ_w^2
Total	$nk-1$	$\sum_{i=1}^{n}\sum_{j=1}^{k}(R_{ij} - \overline{R}_{..})^2$	

Passo 3: Estimar a distribuição do consumo habitual de acordo com a variação intrapessoal.

Se a variância intrapessoal é σ_w^2, então a média das variâncias de n recordatórios independentes para um indivíduo é $\frac{\sigma_w^2}{n}$. Se a variância interpessoal, ou seja, a variância da distribuição do consumo habitual é σ_b^2, segue que a distribuição empírica da média intrapessoal tem variância igual a $\sigma_b^2 + \frac{\sigma_w^2}{n}$. Com isto, constrói-se

um conjunto de valores intermediários com variância σ_b^2 para reduzir cada média individual em relação à média global, ou seja:

[III] valor intermediário= $(1-w)$(média global) + w(média individual),

em que, o fator de redução, é a raiz quadrada da razão da variância interpessoal para a variância da distribuição das médias intrapessoais, isto é:

[IV] $w = \sqrt{\frac{\sigma_b^2}{\sigma_b^2 + \frac{\sigma_w^2}{n}}}$

Por contrução, a distribuição empírica dos valores intermediários tem uma média igual a média global dos R24h. A porcentagem do grupo com consumo menor do que algum valor recomendado para um determinado nutriente, é estimado pela porcentagem dos valores intermediários que são menores do que esse valor. Embora os valores intermediários [III] sejam baseados na média individual, o seu propósito é apenas descrever a distribuição do consumo habitual.

Se os desvios intra e interpessoal são normalmente distribuídos, a distribuição do consumo habitual é descrita apenas pela média e variâncias. Este caso é raro, em parte porque o consumo (habitual ou de um único dia) é por definição não negativo. Existem situações, como no caso de alguns indivíduos que podem ter consumo habitual duas vezes maior que o consumo habitual do grupo, mas um indivíduo não pode ter um consumo habitual regulamente igual ao consumo do grupo. Similarmente, a magnitude dos desvios do consumo que excedem o consumo individual pode ser maior que os desvios daqueles que tem um consumo menor. Portanto, a distribuição do consumo observado tende ser viesada à direita, ou seja, podem existir um pequeno número de valores grandes. Desse modo, a distribuição do consumo não exibirá a simetria correspondente a uma distribuição normal em relação a média.

Considerando esta situação, frequentemente é necessário utilizar uma transformação para os dados. Por exemplo, se os dados tem uma distribuição altamente viesada, então aplica-se uma transformação logaritmica a cada observação de consumo, obtendo uma aproximação da distribuição normal. Para dados menos viesados, aplicam-se transformações mais fracas, tais como, a raiz quadrada ou cúbica, que são as mais frequentes.

Se uma transformação particular produz dados normalmente distribuídos, a distribuição dos dados não transformados pode ser descrito em termos da distribuição normal e da transformada.

A expressão que relaciona valores na escala transformada do consumo habitual na escala original é chamada de transformação inversa. Para retornar aos dados originais, pode-se fazer duas suposições:

- Suposição A: aplica-se a simples transformação inversa aos dados.

- Suposição B: aplica-se a transformação inversa aos dados com um ajuste para o retorno à escala original.

A maioria dos métodos estudados neste trabalho utilizam a suposição B, para relacionar os valores na escala transformada do consumo habitual na escala original.

Tooze et al. (2006a) apresenta os detalhes de obtenção do fator de ajuste a ser aplicado na transformação inversa, que são mostrados na tabela (2).

Tabela 2. Fator de ajuste do viés a ser aplicado na transformação inversa.

Transformação	g(R)	h(r)	Fator de ajuste do viés
Logaritmo	$r = \log(R)$	$R = \exp(r)$	$\frac{1}{2}\exp(b_i^*)\sigma_w^2$
Power	$r = R^{\frac{1}{\lambda}}$	$R = r^{\lambda}$	$\frac{1}{2}\lambda(\lambda-1)(b_i^*)^{\lambda-2}\sigma_w^2$
Box-Cox	$r = \lambda(R^{\frac{1}{\lambda}} - 1)$	$R = (\frac{r}{\lambda} + 1)^{\lambda}$	$\frac{1}{2}[\frac{(\lambda-1)}{\lambda}][\frac{b_i^*}{\lambda+1}]^{\lambda-2}\sigma_w^2$

A seguir, são apresentados os métodos utilizados na análise de dados de consumo baseados no modelo com erro de medida.

1.4 Método do Iowa State University - ISU

Nusser et al. (1996a) descreveram um método desenvolvido na *Iowa State University* para modelar o consumo habitual. O método ISU é baseado na suposição B, no qual um R24h é não viesado para o consumo habitual na escala

original.

Esse método possibilita ajustar os dados observados dos R24h para um indivíduo qualquer sujeito a vieses, tais como, o tempo, os dias da semana e efeitos da amostra de tempo em tempo.

O método ISU é baseado em um modelo complexo que usa a transformação em dois estágios para obter os dados do R24h aproximadamente normal.

Este método permite que a variância intrapessoal varie entre os indivíduos, levando em conta o fato de que alguns indivíduos tendem a ter uma dieta mais variada que outros.

Sob a suposição B, simplesmente aplicar a transformação inversa pode produzir um viés na distribuição do consumo habitual. Então, o método ISU incorpora o ajuste para este viés.

O método ISU pode também levar em conta viés devido à sazonalidade, dias da semana como parte do ajuste.

Este método é usualmente indicado para a obtenção da distribuição de consumo de nutrientes numa população.

1.5 Método do Iowa State University Foods - ISUF

O método ISU foi desenvolvido para modelar a distribuição do consumo habitual. Este é o caso utilizado para a maioria dos nutrientes, e alguns grupos de alimentos comumente consumidos. Contudo, para alimentos, grupos de alimentos e nutrientes consumidos esporadicamente, é possível observar o não consumo em um particular dia. Isto pode gerar uma distribuição com um excesso de zeros na cauda esquerda, sendo difícil obter uma normalização desses dados através de uma transformação.

Nusser et al. (1996b) propuseram então, um método para modelar o consumo habitual que leva em conta esse número excessivo de zeros. As observações nulas são tratadas separadamente das observações positivas. Esta separação é proposta considerando que a média intrapessoal de n dias seja dada por:

$$\frac{\text{consumo total do alimento}}{n} = \frac{k}{n}\frac{\text{consumo total do alimento}}{k},$$

em que k é o número de dias no qual o consumo alimentar é observado, ou seja, o número de R24h realizados (consumo em dias). Para valores grandes de n, a equação anterior expressa dois produtos do consumo médio a longo prazo como o produto de dois componentes. A primeira componente, $(\frac{K}{n})$, corresponde a probabilidade do consumo estimado num dado dia. A segunda componente, $\frac{\text{consumo total do alimento}}{k}$, corresponde a média a longo prazo da quantidade consumida em determinados dias.

A primeira etapa do método (ISUF) estima a distribuição de probabilidade de consumo num dado dia baseado na freqüência relativa dos dados dos R24h observados sem levar em conta as observações nulas. Aplica-se também, parte do conjunto de zeros existentes nos R24h e ajusta-se o consumo diário para um indivíduo qualquer levando-se em conta os viéses, tais como tempo, dias da semana e efeitos das amostras de tempo em tempo, dentre outros. Cada valor da proporção estimada de indivíduos na população é calibrado e observado através de contagem do número de indivíduos que consomem de $0, \ldots, n$ vezes os alimentos registrados nos R24h.

Na segunda etapa, a distribuição do consumo é estimada pelo método ISU aos dados dos R24h eliminando-se os zeros. O Método *Iowa State University* assume a suposição B, que considera que o R24h é não viesado para o consumo habitual diário na escala original.

A modelagem estatística para o consumo episódico de alguns alimentos enfrentam alguns desafios como:

a) Relatar os dias sem consumir um particular alimento ou grupo de alimento;

b) Levar em conta os dados de quantidade de consumo em determinado dia que são geralmente positivamente enviesados e tem valores extremos na cauda superior da distribuição do consumo;

c) Distinguir a variabilidade intrapessoal, que consiste da variação dia-

a-dia no consumo e erros aleatórios da variação interpessoal;

d) Permitir a correlação entre a probabilidade do consumo e a quantidade de consumo de um determinado alimento;

e) Relacionar a informação da covariável (sexo, idade, raça, nível educacional e outras) para o consumo habitual.

O método desenvolvido pelo *Iowa State University* para estimar o consumo esporádico de alimentos (ISUF), reune os três primeiros itens acima, que consiste em relatar os dias sem consumo de alimentos ou grupos de alimentos, permitindo a quantidade de consumo em determinado dia.

Finalmente, o método ISUF combina a distribuição da probabilidade de consumo com o consumo-dia para obter a distribuição estimada do consumo habitual, apesar de tratar essa combinação como modelos independentes.

1.6 Método National Cancer Institute - NCI

O Método do *National Cancer Institute* (NCI) é uma ampliação do método ISUF, pois este, não permite a correlação entre a probabilidade de consumo e a quantidade consumida e não inclui as covariáveis ao modelo. Desse modo, o *National Cancer Institute* desenvolveu este método, proposto por Tooze et al. (2006b). No Método do *National Cancer Institute* (NCI), supõe-se que o R24h é um instrumento não viesado do consumo alimentar habitual esporádico para alimentos consumidos.

Retornando ao modelo com erro de medida clássico, tem-se que R_{ij} e T_{ij} como sendo o consumo verdadeiro de um nutriente ou alimento e o correspondente R24h, para o indivíduo i, $i = 1, 2, \ldots, n$, no dia j, $j = 1, 2, \ldots, j_i$, isto é:

$$R_{ij} = T_i + \epsilon_{ij}, \qquad \text{e} \quad E(\epsilon_{ij}|i) = 0, \tag{19}$$

em que $T_i \sim \text{Normal}(\mu_w, \sigma_w^2)$ e $\epsilon_{ij} \sim \text{Normal}(0, \sigma_\epsilon^2)$, $i = 1, 2, \ldots, n$, $j = 1, 2, \ldots, j_i$ e supondo $k = max\,\{j_i\}$, $i = 1, 2, \ldots, n$.

Supondo uma regressão linear de T_i nas covariáveis $X_i = (X_{i1}, \ldots, X_{ik})$, $k = p + q$, tem-se

$$T_i = \beta_0 + \beta_X^t X_i + u_i + \epsilon_{ij}, \qquad \text{e} \quad u_i \sim \text{Normal}(0; \sigma_u^2). \tag{20}$$

Das equações (19) e (20), os consumos seguem o modelo linear com erro de medida da forma:

$$R_{ij} = \beta_0 + \beta_x^t X_i + u_i + \epsilon_{ij}, \tag{21}$$

para $i = 1, 2, \ldots, n$ e $j = 1, 2, \ldots, j_i$, com os parâmetros populacionais com efeito fixo $\beta = (\beta_0, \beta_x^t)^t$, os efeitos aleatórios u_i que representam os desvios do efeito individual em relação a população do consumo habitual definida através de covariáveis X_i, e erro aleatório intrapessoal ϵ_{ij}. Para este caso também, será necessário dois ou mais R24h para obter a variância intra e interpessoal, sendo os parâmetros do modelo igualmente estimados.

Como no método ISUF para o consumo esporádico de alimentos e nutrientes, a primeira parte do modelo estima a probabilidade de consumir determinado alimento (consumo positivo descrito pelo R24h), usando uma Regressão Logística com efeitos aleatórios dado por:

$$Logit(\text{Probabilidade de consumo relatado no R24h}) = Intercepto_1$$

$$+\text{Inclinação}_1.\text{Covariáveis} + \text{efeito individual em relação à população}_1,$$

O intercepto, a inclinação e a variância do efeito individual em relação a população são parâmetros do modelo e o índice 1, indica a primeira parte do modelo.

A segunda parte do modelo consiste em especificar a quantidade consumida de um alimento em determinado dia usando os dados descritos nos R24h. Esta parte do modelo também pode incorporar a informação das covariáveis para estimar a quantidade de consumo. Porém, as covariáveis da segunda parte, não tem que ser necessariamente as mesmas da primeira parte.

Desse modo, a segunda parte do modelo é dada por:

$$\text{Quantidade consumida relatada nos R24h (em escala transformada)} = Intercepto_2 + \text{Inclinação}_2.\text{Covariáveis} + \text{efeito individual em relação à população}_2 + \text{variabilidade intrapessoal}_2.$$

É importante ressaltar que o modelo é específicado na escala transformada para que o efeito individual em relação a população e a variabilidade intrapessoal aleatória possam ser normalmente distribuídos.

O método NCI ajusta as duas partes do modelo simultaneamente. Primeiramente, os dois efeitos individuais em relação a população são modelados como variáveis aleatórias correlacionadas. Em seguida, algumas covariáveis, que podem ser as mesmas em ambas as partes do modelo, induz a correlação entre elas.

O modelo é ajustado pelo Método da Máxima Verossimilhança para encontrar a correlação e todos os parâmetros do modelo são estimados silmultaneamente, usando um modelo de efeitos não linear. Além disso, a transformação Box-Cox é usada para a normalização dos dados e os parâmetros são estimados como parte do procedimento de máxima verossimilhança.

2 OBJETIVO

O objetivo deste trabalho é avaliar as metodologias estatísticas na análise de dados de consumo alimentar e aplicá-los a um conjunto de dados de um levantamento nutricional de idosos do município de Avaré, São Paulo.

3 METODOLOGIA

3.1 Obtenção de dados de consumo nutricional

Os dados utilizados neste trabalho foram de três recordatórios 24 horas obtidos de uma subamostra de um banco de dados de idosos selecionados das Unidades Básicas de Saúde do município de Avaré para avaliar a qualidade e satisfação com a vida (Processo FAPESP n° 2006/06807-0). O tamanho amostral inicialmente calculado foi de 365 idosos, sendo considerada uma prevalência de satisfação com a vida de 0,5 (maior valor de prevalência por ser uma característica desconhecida na população), uma margem de erro amostral de 5%, valor z igual a 1,96 relativo a um intervalo de confiança de 95% bilateral. O sorteio dos participantes da amostra foi feito estratificando por faixa etária, de dez em dez anos.

A subamostra considerada para o presente estudo foi de 20% dos idosos, perfazendo um total de 73 indivíduos de ambos os sexos. Esse tamanho amostral foi proposto uma vez que o estudo objetivava validar um questionário de frequência alimentar para idosos (Processo CNPq n° 402533/2007-0), sendo que, para este tipo de estudo, recomenda-se uma amostra de 50 a 100 participantes Slater et al. (2003). O sorteio dos participantes foi novamente estratificado por faixa etária.

Após a aplicação dos três R24 horas, os dados obtidos foram digitados no programa Nutwin (Programa de Apoio à Nutrição Nutwin, v.1.5, 2002) para a obtenção dos consumos de macros e micronutrientes. Os dados de consumo foram exportados para um arquivo texto e importados em uma planilha Excel.

Além dos dados de consumo, foram obtidos dados referentes a características sociodemográficas, de prevenção e morbidades.

3.2 Aplicação do Metodo ISU

Para analisar dados de consumo de nutrientes e alimentos utilizando os métodos ISU e ISUF, programas e rotinas apropriadas devem ser utilizados e desenvolvidas. Para ambos os métodos, ISU e ISUF pode-se utilizar o programa SIDE, (*Software for Intake Distribution Estimation*) desenvolvido pela *Iowa State University*. Este programa foi inicialmente desenvolvido utilizando o aplicativo IML do programa SAS, que recebeu o nome de SIDE/IML. Após isto, foi desenvolvido para a plataforma LINUX/UNIX, recebendo o nome de SIDE e, posteriormente, a versão para *Windows*, que é o PC-SIDE. A metodologia de construção do PC-SIDE é descrito por Nusser et al. (1996b) e Nusser et al. (1996a).

O PC-SIDE utiliza diferentes metodologias para analisar dados de consumo de nutrientes e de alimentos. Quando trata-se de dados de consumo de nutrientes, o PC-SIDE faz um ajuste preliminar nos dados e, se necessário, aplica-se uma transformação semiparamétrica aos dados de consumo buscando normalizá-los. Em seguida, estima-se as componentes de variância, as variâncias intra e interpessoais do consumo de nutrientes e para estimar a distribuição do consumo habitual do nutriente, faz a transformação inversa aos dados, relacionando à escala original. Para se obter a distribuição estimada do consumo, são necessários pelo menos dois R24h. No caso de apenas um R24h, é possível obter a distribuição de consumo, mas é preciso repetir os dados simulando dois R24h e informar a razão entre os componentes de variâncias inter e intrapessoal, Dodd (2004).

Quando se tem dados de consumo de alimentos, o PC-SIDE estima a distribuição de probabilidade de um indivíduo consumir um alimento em um dado dia. Este método também pode estimar a distribuição do consumo habitual do nutriente, aplicado para consumos positivos de nutrientes e estima a distribuição do consumo habitual para o consumo de um único dia.

Para dados de consumo esporádico de alimentos pode-se utilizar o método ISUF com o programa PC-SIDE, desde que o número de observações nulas não seja muito grande.

A entrada de dados no programa PC-SIDE referente a pelo menos dois R24h deve ser da seguinte forma: primeiramente, quando se trata de dados fornecidos pelos R24h, não é necessário que número observações para cada indivíduos seja a mesma, isto é, como o programa utiliza um modelo com erros de medidas para obtenção das variâncias inter e intrapessoais, não é necessário ter o mesmo número de R24h para todos os indivíduos. Para alimentos, se existem alimentos que são frequentemente consumidos, então, o número de observações para cada indivíduo também não precisa ser o mesmo para todos os indivíduos no conjunto de dados fornecido pelo R24h. Desse modo, não pode haver consumo esporádico de nutrientes relatados em algum dos R24h.

Os dados devem estar em arquivo texto ASCII, com uma observação em cada linha e uma única variável indicadora de R24h e outras informações para cada indivíduo. O arquivo deve ser ordenado, considerando os R24h para cada indivíduo.

Para fazer a análise dos dados é necessário definir os tipos variáveis consideradas. O programa PC-SIDE, oferece alguns tipos de variáveis, tais como, *Nutrient*, *Food*, *Individual ID*, *Sample day*, *Weight*, *Class indy*, *Class obs*, *Continuous indv*, *Continuous obs* e *Nonanalysis*.

De acordo com os dados fornecidos ao programa, a execução segue os seguintes passos:

Passo 1: Uma transformação Box-Cox é obtida para normalizar os dados. Caso não seja possível obter a normalização, o programa encerra a análise.

Passo 2: Após a normalização dos dados, quando possível, o programa utiliza um modelo com erros de medida para obter as variâncias inter e intrapessoais. Essas estimativas são utilizados para obter a distribuição do consumo habitual em escala normal. Alguns efeitos do modelo podem ser considerados como dia da semana, sazonalidade, dentre outros. O programa não aceita o uso de covariáveis.

Passo 3: Após o ajuste de distribuição do consumo, uma transformação inversa é aplicada afim de retornar à escala original. Neste passo, é possível

estimar a probabilidade de inadequação de consumo de determinado nutriente ou alimento, ou seja, a probabilidade da população consumir acima ou abaixo de um determinado padrão. Esses valores são comumente obtidos das tabelas de recomendações, chamadas DRI (2006).

3.3 Aplicação do Método NCI

O método NCI, desenvolvido para estimar a distribuição do consumo de nutrientes, com um consumo diário e esporádico, é uma versão ampliada do método ISU, em que, baseado num modelo com erro de medida pode-se incluir covariáveis.

O NCI disponibiliza as macros *Mixtran* e *Distrib* do *Software* SAS, versão 9.1 para estimação da distribuição do consumo habitual de nutrientes e alimentos, estendendo a consumos esporádicos. A macro *Mixtran* pode ser usada para estimar os efeitos individuais das covariáveis no consumo de alimentos ou de nutrientes e estima todos os parâmetros do modelo. A *Mixtran* ajusta um modelo com efeitos não linear para dois ou mais R24h, usando o procedimento NLMIXED do SAS. Para o consumo esporádico de alimentos, a macro ajusta um modelo em duas partes: um modelo logístico, indicando o consumo ou não do nutriente ou alimento, e um modelo com erro de medida, levando em conta a quantidade consumida daquele nutriente ou alimento.

A sintaxe da macro *Mixtran* é:

```
%mixtran(data=, response=, foodtype=, subject=, repeat=,
covars_prob=, covars_amt=,
outlib=, modeltype=, lambda=, replicate_var=, seq=,
weekend=, vargroup=,
numvargroups=, subgroup=, start_val1=, start_val2=, start_val3=,
vcontrol=, nloptions=, titles=, printlevel=);,
```

em que, os parâmetros utilizados são: *data*, *response*, *foodtype*, *subject*, *repeat*, *covars_prob*, *covars_amt*, *outlib*, *modeltype*, *lambda*, *replicate_var*, *seq*, *weekend*, *var-*

group, *numvargroups*, *subgroup*, *start_val*1, *start_val*2, *start_val*3, *vcontrol*, *nloptions*, *titles*, *printlevel*.

O ajuste final é dado pela multiplicação dos dois modelos, obtendo-se a distribuição do consumo habitual, incluindo as covariáveis de interesse, que podem ser as mesmas tanto na primeira parte quando na segunda parte, que é feita pela macro *Distrib*, cuja sintaxe é:

```
%Distrib (seed=, nsim_mc=, modeltype=, pred=, param=,
outlib=, cutpoints=, ncutpnt=, byvar=, subgroup=,
subject=, titles=, food=);,
```

em que, os parâmetros utilizados são: *seed*, *nsim_mc*, *modeltype*, *pred*, *param*, *outlib*, *cutpoints*, *ncutpnt*, *subgroup subject*, *titles*, *food*.

4 RESULTADOS E DISCUSSÃO

4.1 Análise dos dados do consumo

Os dados de consumo foram obtidos de uma subamostra de 73 idosos do município de Avaré, que fizeram parte de uma pesquisa sobre qualidade de vida na terceira idade. Algumas características desses idosos são mostrados na tabela 3.

Tabela 3. Descrição da amostra de idosos de Avaré, 2009

Variáveis	*N*	%
Sexo		
Masculino	32	43,84
Feminino	41	56,16
Etnia		
Branco	69	94,52
Negro	3	4,11
Amarelo	1	1,37
Estado Nutricional		
Desnutrido	9	12,64
Eutrófico	36	50,70
Sobrepeso	26	36,62
Dentição		
Sim	147	84,48
Não	27	15,52
Faz uso de suplemento		
Sim	13	18,06
Não	59	81,94
Tabagismo		
Sim	8	10,96
Não	65	89,04
Atividade Física		
Sim	26	35,62
Não	47	64,38

Considerando o consumo dos nutrientes, cálcio, zinco, magnésio, vitamina A, vitamina B12, vitamina C e vitamina E, a tabela 4 mostra uma descrição desses consumos, considerando os três R24h.

Tabela 4. Descrição do consumo de nutrientes da amostra de idosos. Avaré, 2009

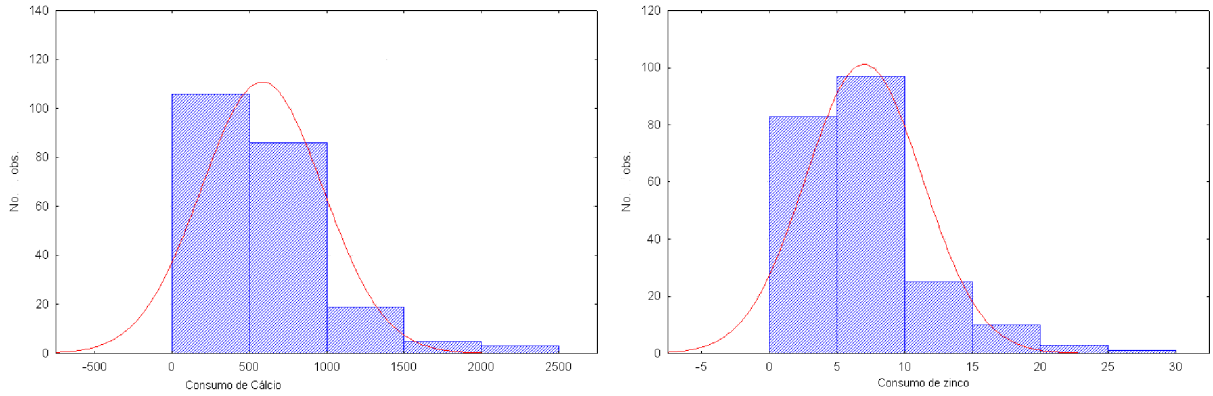
Variáveis	N	Média	Mediana	Mínimo	Máximo	Variância	Desvio Padrão
Idade	219	71,51	70,00	60,00	87,00	41,56	6,45
Altura	216	1,64	1,65	1,43	1,85	0,01	0,09
Peso	213	71,43	72,00	39,00	109,00	182,39	13,51
IMC	213	26,60	25,86	16,66	42,58	24,14	4,91
Cálcio	219	583,57	507,15	40,21	2280,17	155436,71	394,25
Magnésio	219	185,62	165,77	29,56	649,83	10164,64	100,82
Vitamina A	219	534,26	284,30	4,14	5955,29	578690,02	760,72
Vitamina B12	219	2,63	2,00	0,09	14,44	4,37	2,09
Vitamina C	219	149,12	42,98	0,33	2045,55	73782,90	271,63
Vitamina E	219	9,18	7,24	0,95	45,67	37,10	6,09
Zinco	219	6,99	5,92	0,94	28,36	18,65	4,32

A tabela 5 mostra as estatísticas descritivas do consumo de nutrientes estratificado por sexo.

Tabela 5. Descrição do consumo de nutrientes da amostra de idosos estratificado por sexo. Avaré, 2009

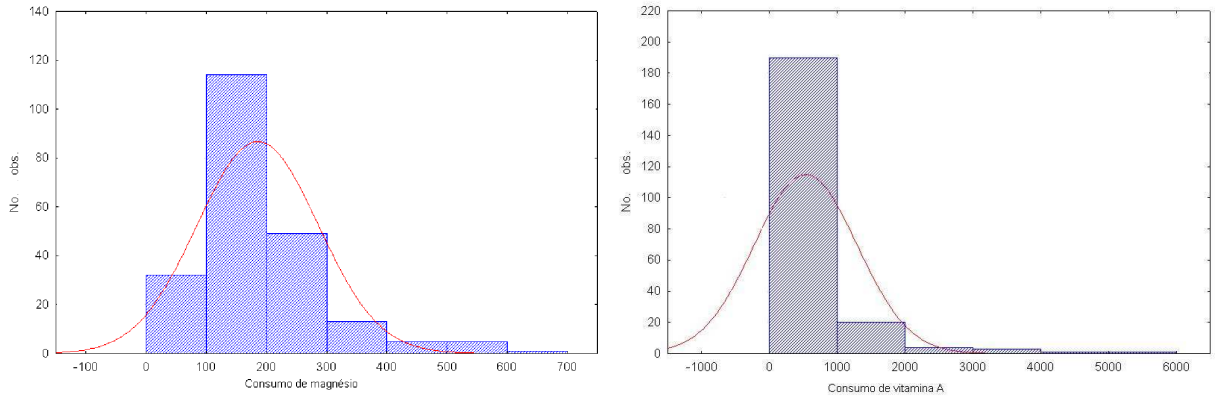
Sexo	Variáveis	N	Média	Mediana	Mínimo	Máximo	Variância	Desvio Padrão
Masculino	Idade	96	72,75	70,50	63,00	87,00	43,26	6,58
	Altura	96	1,70	1,70	1,60	1,85	0,00	0,06
	Peso	93	76,44	76,00	56,00	109,00	156,70	12,52
	IMC	93	26,45	26,12	18,52	42,58	21,14	4,60
	Cálcio	96	619,46	576,16	104,51	2280,17	173833,10	416,93
	Magnésio	96	225,07	192,41	92,77	649,83	13893,71	117,87
	Vitamina A	96	602,46	310,77	4,14	5955,29	743867,99	862,48
	Vitamina B12	96	3,04	2,36	0,11	14,44	5,83	2,41
	Vitamina C	96	100,13	41,67	1,13	883,06	25526,26	159,77
	Vitamina E	96	10,36	8,14	2,66	45,67	47,77	6,91
	Zinco	96	8,28	7,22	1,83	28,36	24,06	4,91
Feminino	Idade	123	70,54	70,00	60,00	80,00	38,41	6,20
	Altura	120	1,59	1,60	1,43	1,75	0,01	0,08
	Peso	120	67,56	68,00	39,00	96,00	169,08	13,00
	IMC	120	26,72	25,00	16,66	37,50	26,62	5,16
	Cálcio	123	555,56	469,48	40,21	2114,70	140581,12	374,94
	Magnésio	123	154,83	141,44	29,56	444,04	5163,92	71,86
	Vitamina A	123	481,03	258,75	10,64	4551,80	448294,66	669,55
	Vitamina B12	123	2,31	1,95	0,09	9,43	3,03	1,74
	Vitamina C	123	187,35	45,21	0,33	2045,55	108601,95	329,55
	Vitamina E	123	8,25	6,96	0,95	31,94	27,13	5,21
	Zinco	123	5,98	5,26	0,94	19,67	12,26	3,50

A distribuição do consumo de cálcio, zinco, magnésio, vitamina A, vitamina B12, vitamina C, vitamina E sem correção pela variabilidade intrapessoal é dada pelas figuras 2 a 5.



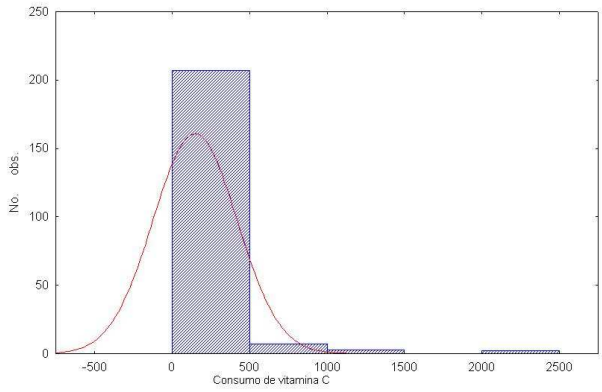
a) b)

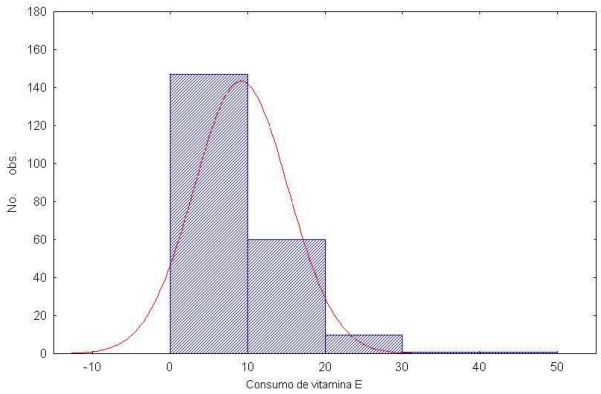
Figura 2 - Distribuição do consumo de (a) cálcio e (b) zinco para os idosos. Avaré, 2009



c) d)

Figura 3 - Distribuição do consumo de (c) magnésio e (d) vitamina A para os idosos. Avaré, 2009





e) f)

Figura 4 - Distribuição do consumo de (e) vitamina C e (f) vitamina E para os idosos. Avaré, 2009

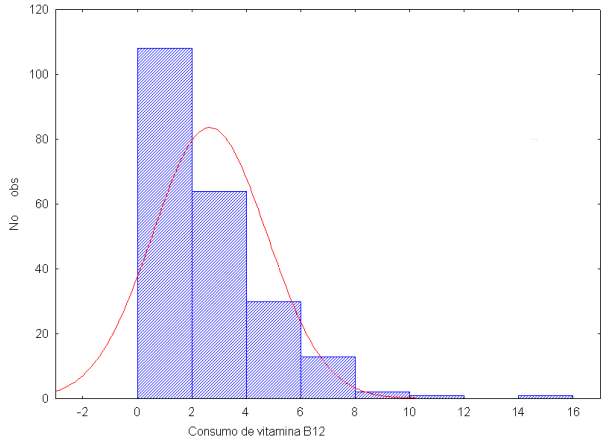
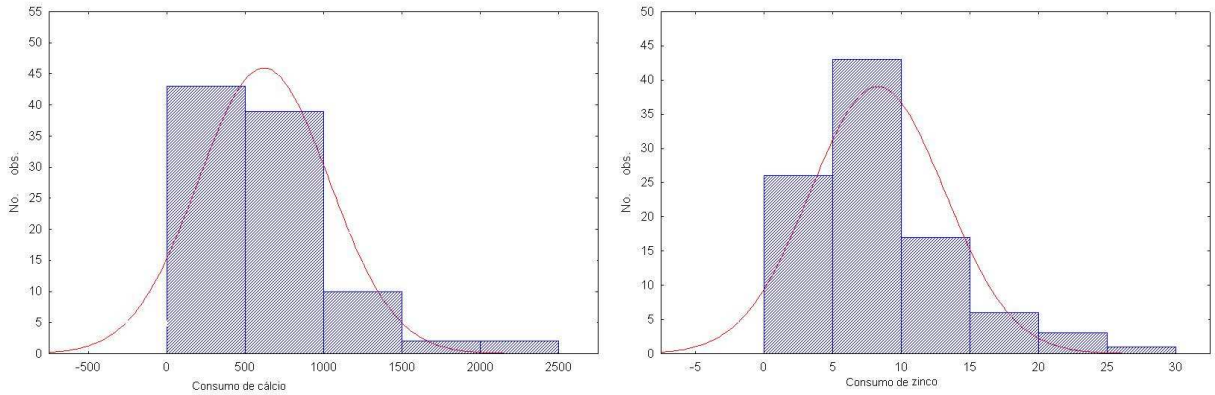


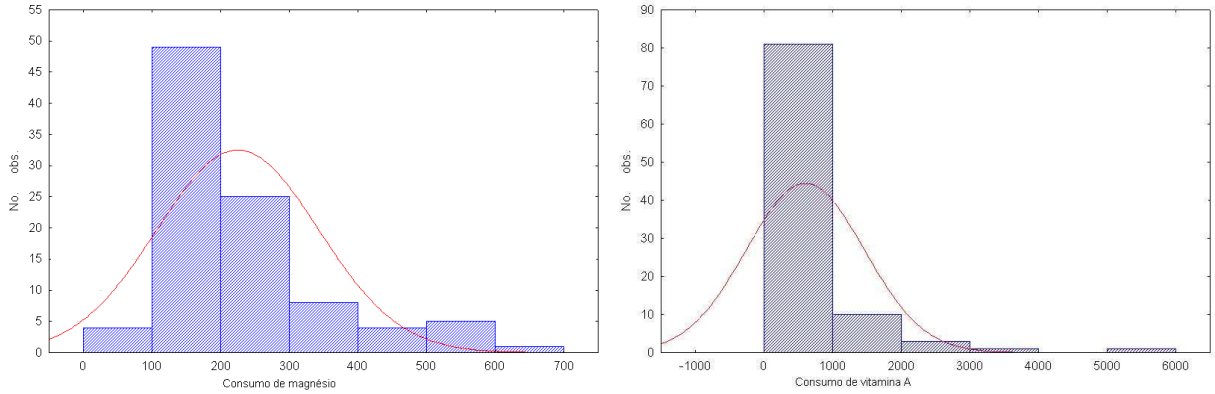
Figura 5 - Distribuição do consumo de vitamina B12 para os idosos. Avaré, 2009

A distribuição do consumo de cálcio, zinco, magnésio, vitamina A, vitamina B12, vitamina C, vitamina E sem correção pela variabilidade intrapessoal é dada pelas figuras a seguir, estratificado por sexo. As figuras de 6 a 9, representam a distribuição do consumo destes nutrientes para o sexo masculino.



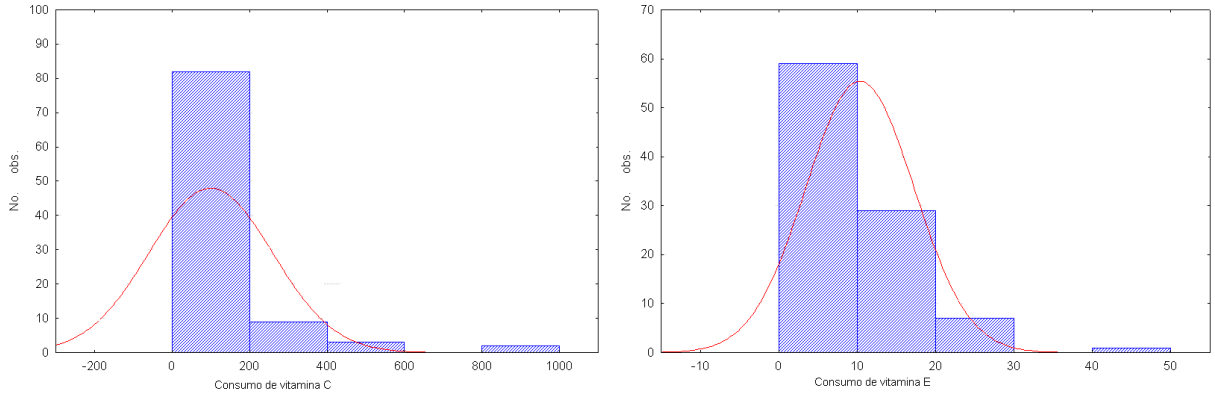
a) b)

Figura 6 - Distribuição do consumo de (a) cálcio e (b) zinco para os idosos do sexo masculino. Avaré, 2009



c) d)

Figura 7 - Distribuição do consumo de (c) magnésio e (d) vitamina A para os idosos do sexo masculino. Avaré, 2009



e) f)

Figura 8 - Distribuição do consumo (e) vitamina C e (f) vitamina E para os idosos do sexo masculino. Avaré, 2009

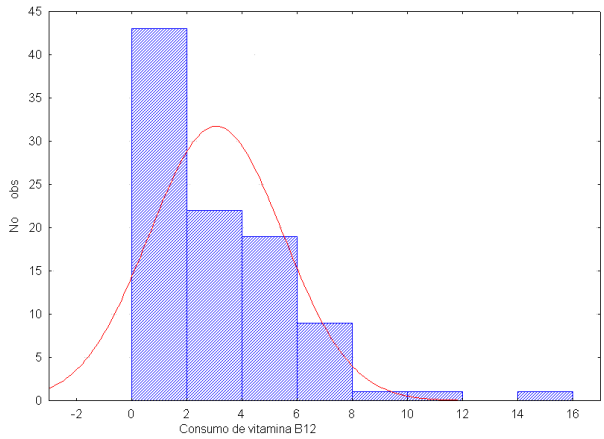
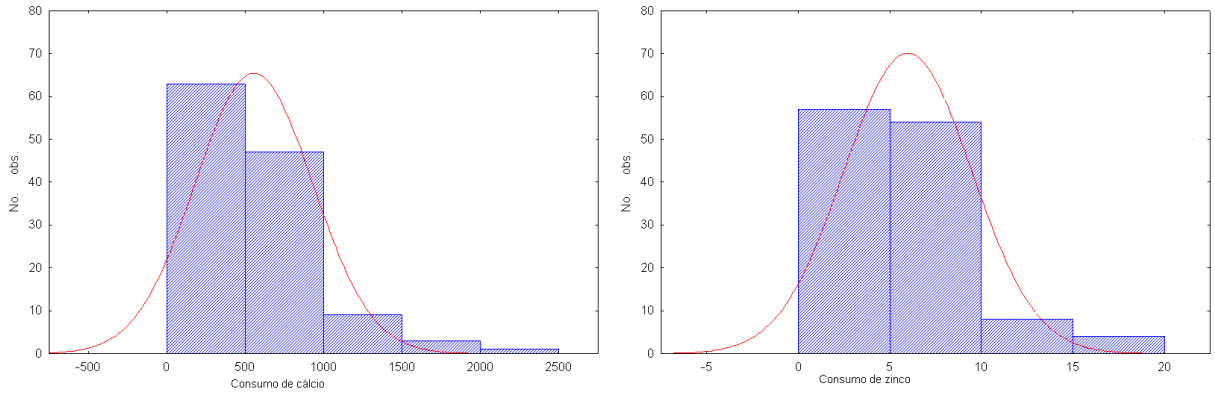


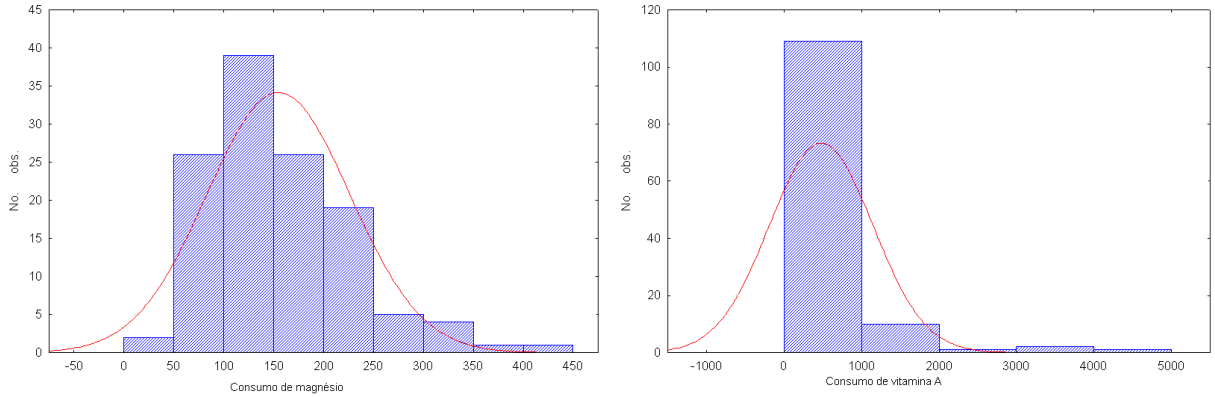
Figura 9 - Distribuição do consumo de vitamina B12 para os idosos do sexo masculino. Avaré, 2009

A distribuição do consumo de cálcio, zinco, magnésio, vitamina A, vitamina B12, vitamina C, vitamina E sem correção pela variabilidade intrapessoal é dada pelas figuras a seguir, estratificado por sexo. As figuras 10 a 13, representam a distribuição do consumo destes nutrientes para o sexo feminino.



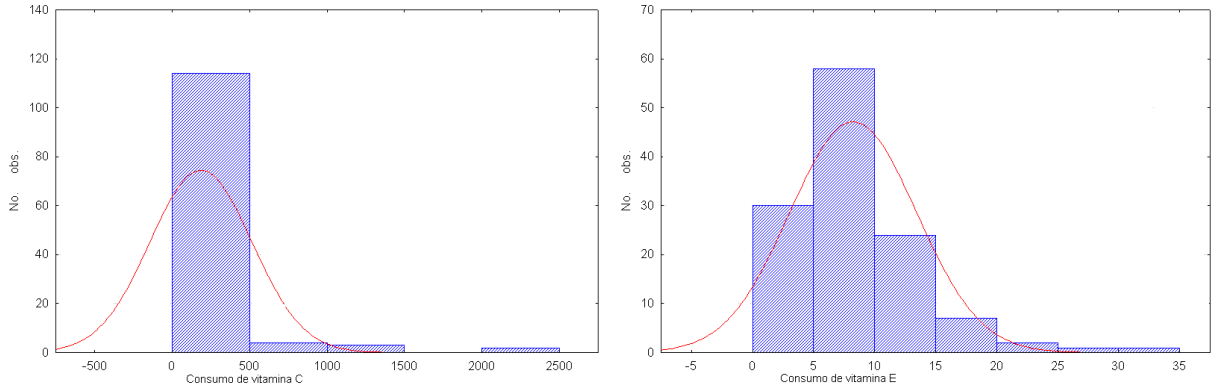
a) b)

Figura 10 - Distribuição do consumo de (a) cálcio e (b) zinco para os idosos do sexo feminino. Avaré, 2009



c) d)

Figura 11 - Distribuição do consumo de (c) magnésio e (d) vitamina A para os idosos do sexo feminino. Avaré, 2009



e) f)

Figura 12 - Distribuição do consumo de (e) vitamina C e (f) vitamina E para os idosos do sexo feminino. Avaré, 2009

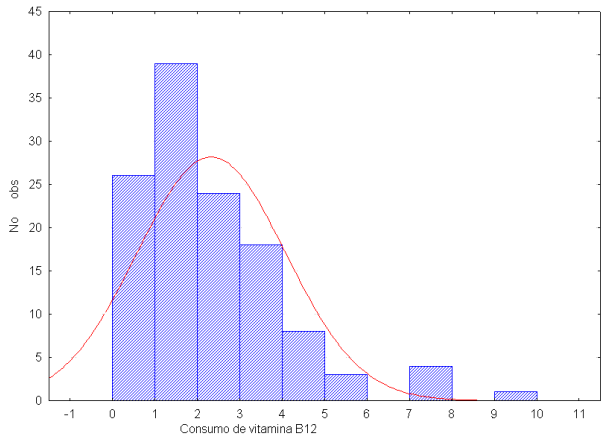


Figura 13 - Distribuição do consumo de vitamina B12 para os idosos do sexo feminino. Avaré, 2009

4.2 Resultados da análise dos dados do consumo

Para avaliar a adequação da dieta em relação a algum padrão recomendado é preciso obter a distribuição do consumo habitual. Desse modo, são necessários métodos apropriados para estimar a ingestão de alimentos e nutrientes de grupos populacionais. Dentre os nutrientes avaliados, serão analisados através da aplicação dos métodos citados, obtidos dos três R24h.

Inicialmente, utilizando o programa PC-SIDE, v.1.0 e utilizando as rotinas do programa SAS, (Tooze et al. (2006a)) a distribuição do consumo habitual para os nutrientes citados é feita considerando um modelo com erros de medida. Desse modo, as probabilidades acima ou abaixo de um valor recomendado são apresentados na tabela 6.

Tabela 6. Probabilidade de inadequação de nutrientes para idosos de ambos os sexos pelos métodos ISU e NCI. Avaré, 2009

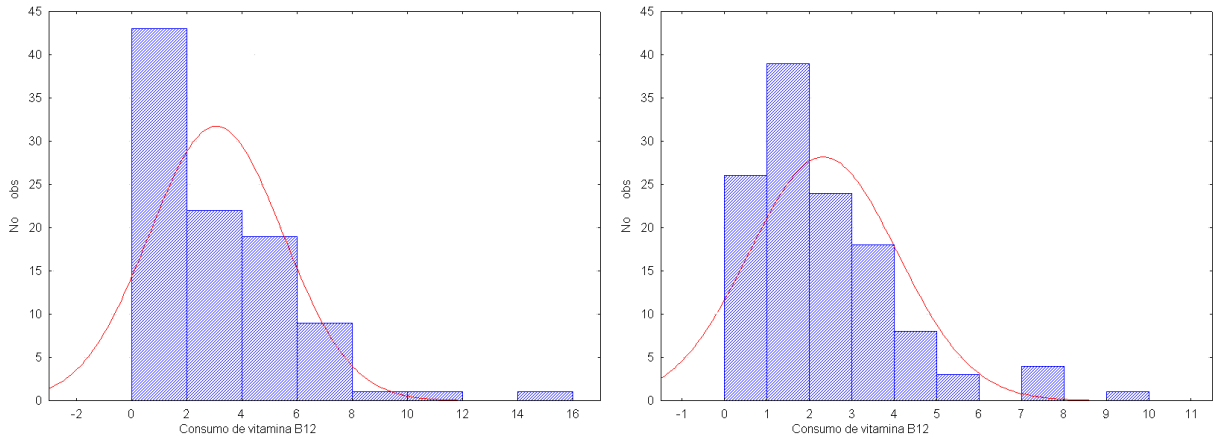
		ISU	NCI
Nutriente	Recomendação	Prob.abaixo	Prob.abaixo
Cálcio			
Masculino	1200	0,94	0,94
Feminino	1200	0,94	0,95
Magnésio			
Masculino	350	0,90	0,95
Feminino	265	0,93	0,97
Zinco			
Masculino	9,4	0,82	0,72
Feminino	6,8	0,84	0,73
Vitamina A			
Masculino	625	0,58	0,69
Feminino	500	0,62	0,70
Vitamina B12			
Masculino	1,2	0,32	0,27
Feminino	1,2	0,44	0,34
Vitamina C			
Masculino	75	0,58	0,60
Feminino	60	0,19	0,32
Vitamina E			
Masculino	12	0,93	0,74
Feminino	12	0,84	0,94

Os métodos utilizados se adequaram aos nutrientes considerados. Pelo método ISU, foi possível através do PC-SIDE normalizar os dados do consumo de nutrientes e obter a distribuição desses consumos bem como estimar a probabilidade de inadequação (Tabela 6) com exceção do consumo de vitamina B12. Essa probabilidade é obtida corrigindo os dados pela variabilidade intrapessoal através de um modelo com erro de medida.

No caso do método NCI, além da distribuição do consumo dos nutrientes considerados, foi possível incluir as covariáveis sociodemográficas, de prevenção e morbidades referidas.

4.2.1 Análise dos dados do consumo de vitamina B12 para ambos os sexos.

As figuras 14 mostram a distribuição do consumo de vitamina B12 para o sexo masculino e feminino, respectivamente.

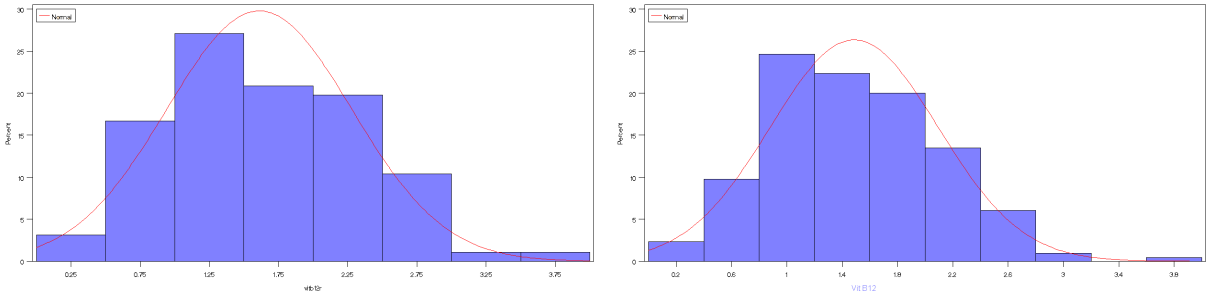


a) b)

Figura 14 - Distribuição do consumo de vitamina B12 para os idosos (a) do sexo masculino e (b) do sexo feminino. Avaré, 2009

Como a distribuição é bastante assimétrica, não foi possível obter a probabilidade de inadequação pelo PC-SIDE, sendo aplicado o método ISU manualmente a esses dados.

Afim de obter uma distribuição mais próxima à distribuição normal, algumas transformações foram testadas. A que ficou um pouco mais simétrica foi quando os dados foram transformados aplicando-se a raiz quadrada. As figuras 15 mostram o histograma e o ajuste de distribuição normal após a transformação.



a) b)

Figura 15 - Distribuição do consumo após a transformação raiz quadrada de vitamina B12 para idosos (a) do sexo masculino e (b) do sexo feminino. Avaré, 2009

Após a transformação, foi aplicado com procedimento de ANOVA para obtenção das variabilidades inter e intrapessoais.

Tabela 7. ANOVA para o consumo de vitamina B12 pelos idosos do sexo masculino.

F. Variação	GL	SQ	QM
Interpessoal	30	21,55	0,72
Intrapessoal	62	18,41	0,29
Total	92	39,96	

De acordo com a tabela 7, tem-se que as estimativas das variâncias inter e intrapessoais são:

$s_w^2 = 0,29,$

$s_b^2 = 0,72 - \frac{0,29}{3},$

$s_b^2 = 0,62,$

daí, a razão entre a variabilidade intrapessoal e a total é dada por:

$$r = \frac{S_b}{S_{total}} = \sqrt{\frac{1}{1 + \frac{s_w^2}{k(s_b^2)}}} = \sqrt{\frac{1}{1 + \frac{0,29}{3*0,62}}} = 0,97.$$

Para se determinar a média dos k dias de inquérito, eliminando a variabilidade intrapessoal, ajusta-se a média de energia e nutriente de cada indivíduo, ou seja:

$$\text{Valor de vitamina B12 ajustado} = \text{média de vitamina B12} +$$

$$(\text{consumo de vitamina B12} - \text{média de vitamina B12})r.$$

Obtida o valor da variável ajustada, aplica-se a transformação inversa e calcula-se a probabilidade de inadequação que é dada por:

$$z = \frac{(\text{valor recomendado} - \text{média geral ajustada para cada indivíduo})}{\text{Dp}}.$$

Sendo, a recomendação de vitamina B12 para o sexo masculino de 2 mcg, tem-se que:

$$z = \frac{(2 - 2,74)}{1,56} = -0,47,$$

e, de acordo com a distribuição normal padrão, $z = -0,47$ fornece uma probabilidade de inadequação de $p = 0,32$, ou seja, 32% de inadequação.

Agora, considere o consumo de vitamina B12 para o sexo feminino, seguindo os passos da análise anterior. A ANOVA para o consumo de vitamina B12 para o sexo feminino é dada pela tabela 8.

Tabela 8. ANOVA para o consumo de vitamina B12 para o sexo feminino.

F. Variação	GL	SQ	QM
Interpessoal	40	20,97	0,52
Intrapessoal	82	16,50	0,20
Total	122	37,47	

De acordo com a tabela 8, tem-se que as estimativas das variâncias inter e intrapessoais são:

$s_w^2 = 0,20,$

$s_b^2 = 0,52 - \frac{0,20}{3},$

$s_b^2 = 0,46,$

daí, a razão entre a variabilidade intrapessoal e a total é dada por:

$$r = \frac{S_b}{S_{total}} = \sqrt{\frac{1}{1 + \frac{s_w^2}{k(s_b^2)}}} = \sqrt{\frac{1}{1 + \frac{0,20}{3*0,46}}} = 0,98.$$

Analogamente, considerando a recomendação de vitamina B12 para o sexo feminino de 2 mcg, tem-se que:

$$z = \frac{(2 - 2,17)}{1,19} = -0,14,$$

e, de acordo com a distribuição normal padrão, $z = -0,14$ fornece uma probabilidade de inadequação de $p = 0,44$, ou seja, 44% de inadequação.

4.3 Resultados das análises dos dados do consumo de nutrientes utilizando o método NCI e incluindo as covariáveis no modelo

O método proposto por Tooze et al. (2006a), permite a correlação entre a probabilidade de consumo e a quantidade consumida e ainda, inclui covariáveis ao modelo no qual permite observar qual será a influência dessas covariáveis no consumo

habitual de um grupo ou uma população, levando-se em conta um valor recomendado e considerando o sexo. As covariáveis inseridas no modelo são: a idade, índice de massa corpórea, estado nutricional (eutrófico e sobrepeso), se faz uso de suplementos alimentares, se possui problemas intestinais, se é tabagista, uso de prótese e se prática alguma atividade física.

A tabela 9, mostra o ajuste do modelo proposto pelo método NCI do consumo de cálcio para os idosos de ambos os sexos.

Tabela 9. Ajuste do modelo com erro de medida proposto pelo método NCI para o consumo de cálcio estratificado por sexo. Avaré, 2009

	Cálcio			
Sexo	Covariáveis	Estimativas	Erro Padrão	p
Masculino	Intercepto	26,50	21,18	0,22
	Idade	-0,12	0,12	0,34
	Índice de massa corpórea	-0.21	0,23	0,37
	Eutrofíco	-0,84	1,49	0,58
	Sobrepeso	0,24	1,69	0,89
	Uso de prótese	-0,01	0,76	0,98
	Faz uso de suplemento	-0,04	1,85	0,98
	Possui problemas intestinais	-0,92	0,98	0,36
	Tabagismo	0,06	0,92	0,95
	Faz atividade física	0,25	0,87	0,78
	Transformação Box-Cox(λ)	0,15	0,14	0,30
	Variância interpessoal	-0,08	0,88	0,92
	Variância intrapessoal	0,26	0,89	0,77
Feminino	Intercepto	15,14	14,15	0,29
	Idade	-0,17	0,22	0,45
	Índice de massa corpórea	-0,05	0,15	0,76
	Eutrofíco	0,34	1,33	0,80
	Sobrepeso	-1,35	2,93	0,65
	Uso de prótese	3,17	4,09	0,44
	Faz uso de suplemento	-0,17	0,81	0,83
	Possui problemas intestinais	-0,17	0,44	0,70
	Tabagismo	4,12	5,20	0,43
	Faz atividade física	-0,20	0,77	0,79
	Transformação Box-Cox(λ)	0,19	0,19	0,33
	Variância interpessoal	0,14	1,22	0,91
	Variância intrapessoal	0,49	1,18	0,68

Observando a tabela 9, tem-se que não houve influência significativa das covariáveis para o consumo habitual de cálcio para este grupo de idosos tanto do sexo masculino quanto do feminino.

A tabela 10, mostra o ajuste do modelo proposto pelo método NCI do consumo de magnésio para os idosos do sexo masculino e feminino.

Tabela 10. Ajuste do modelo com erro de medida proposto pelo método NCI para o consumo de magnésio estratificado por sexo. Avaré, 2009

	Magnésio			
Sexo	Covariáveis	Estimativas	Erro Padrão	p
Masculino	Intercepto	7,28	1,66	0,00
	Idade	-0,01	0,01	0,49
	Índice de massa corpórea	-0,01	0,03	0,65
	Eutrofíco	0,17	0,28	0,54
	Sobrepeso	0,11	0,37	0,78
	Uso de prótese	-0,02	0,17	0,89
	Faz uso de suplemento	-0,08	0,41	0,84
	Possui problemas intestinais	-0,11	0,11	0,35
	Tabagismo	-0,07	0,20	0,72
	Faz atividade física	-0,18	0,19	0,34
	Transformação Box-Cox(λ)	0,01	-	-
	Variância interpessoal	-1,11	0,10	0,00
	Variância intrapessoal	-1,40	0,24	0,00
Feminino	Intercepto	6,49	5,24	0,22
	Idade	-0,03	0,04	0,47
	Índice de massa corpórea	-0,03	0,05	0,55
	Eutrofíco	0,35	0,54	0,52
	Sobrepeso	0,31	0,61	0,62
	Uso de prótese	0,79	1,05	0,45
	Faz uso de suplemento	-0,01	0,17	0,95
	Possui problemas intestinais	-0,02	0,08	0,84
	Tabagismo	0,76	0,97	0,44
	Faz atividade física	-0,19	0,27	0,49
	Transformação Box-Cox(λ)	0,05	0,25	0,84
	Variância interpessoal	-1,02	1,24	0,42
	Variância intrapessoal	-1,16	1,29	0,37

De acordo com a tabela 10, as covariáveis incluidas no modelo não

tiveram influência no consumo de magnésio dos idosos do sexo masculino e feminino. Houve apenas efeito significativa da variabilidade intrapessoal e interpessoal para o sexo masculino, indicando que não existiu um padrão de consumo para esse nutriente.

A tabela 11 mostra o ajuste do modelo proposto pelo método NCI do consumo de zinco para os idosos do sexo masculino e feminino.

Tabela 11. Ajuste do modelo com erro de medida proposto pelo método NCI para o consumo de zinco estratificado por sexo. Avaré, 2009.

Zinco				
Sexo	Covariáveis	Estimativas	Erro Padrão	p
Masculino	Intercepto	5,45	1,71	0,01
	Idade	-0,03	0,01	0,03
	Índice de massa corpórea	-0,05	0,03	0,11
	Eutrofíco	-0,09	0,29	0,75
	Sobrepeso	-0,23	0,39	0,56
	Uso de prótese	0,08	0,17	0,66
	Faz uso de suplemento	0,00	0,42	0,99
	Possui problemas intestinais	-0,15	0,12	0,20
	Tabagismo	0,27	0,21	0,21
	Faz atividade física	0,18	0,19	0,37
	Transformação Box-Cox(λ)	0,01	-	-
	Variância interpessoal	-0,73	0,10	0,00
	Variância intrapessoal	-1,80	0,62	0,01
Feminino	Intercepto	1,94	0,82	0,02
	Idade	-0,03	0,01	0,00
	Índice de massa corpórea	0,02	0,02	0,34
	Eutrofíco	0,20	0,19	0,29
	Sobrepeso	-0,22	0,32	0,50
	Uso de prótese	0,61	0,29	0,04
	Faz uso de suplemento	-0,11	0,12	0,36
	Possui problemas intestinais	0,00	0,05	0,98
	Tabagismo	0,71	0,33	0,04
	Faz atividade física	-0,20	0,11	0,07
	Transformação Box-Cox(λ)	0,01	-	-
	Variância interpessoal	-0,87	0,09	0,00
	Variância intrapessoal	-2,43	1,27	0,06

Pela tabela 11 pode-se verificar que a idade e as variabilidade intra e

interpessoal tiveram influência significativa no consumo de zinco para os idosos do sexo masculino, enquanto que, a idade, uso de prótese, o tabagismo e a variabilidade interpessoal tiveram influência significava no consumo zinco para o sexo feminino. Pode-se considerar que para ambos os sexos não existiu um padrão de consumo de zinco.

A tabela 12 mostra o ajuste do modelo proposto pelo método NCI do consumo de vitamina A para os idosos do sexo masculino e feminino.

Tabela 12. Ajuste do modelo com erro de medida proposto pelo método NCI para o consumo de vitamina A estratificado por sexo. Avaré, 2009.

Vitamina A				
Sexo	Covariáveis	Estimativas	Erro Padrão	p
Masculino	Intercepto	11,40	6,29	0,08
	Idade	-0,01	0,04	0,78
	Índice de massa corpórea	0,00	0,08	0,98
	Eutrofíco	-0,41	0,78	0,60
	Sobrepeso	-0,32	1,03	0,76
	Uso de prótese	0,17	0,46	0,71
	Faz uso de suplemento	-0,59	1,17	0,62
	Possui problemas intestinais	0,10	0,32	0,76
	Tabagismo	-0,87	0,74	0,25
	Faz atividade física	-1,04	0,75	0,18
	Transformação Box-Cox(λ)	0,03	0,08	0,70
	Variância interpessoal	0,27	0,51	0,60
	Variância intrapessoal	-0,96	0,93	0,31
Feminino	Intercepto	7,42	3,09	0,02
	Idade	-0,06	0,04	0,09
	Índice de massa corpórea	-0,02	0,07	0,82
	Eutrofíco	0,06	0,60	0,92
	Sobrepeso	-0,50	1,03	0,63
	Uso de prótese	0,37	0,94	0,69
	Faz uso de suplemento	-0,17	0,37	0,64
	Possui problemas intestinais	0,01	0,17	0,94
	Tabagismo	2,46	1,36	0,08
	Faz atividade física	-0,19	0,35	0,60
	Transformação Box-Cox(λ)	0,06	0,06	0,37
	Variância interpessoal	0,31	0,39	0,43
	Variância intrapessoal	-1,40	1,65	0,40

Pela tabela 12 pode-se observar que as covariáveis não tiveram influência sobre o consumo de vitamina A para os idosos do sexo masculino e feminino. Apesar de a tabela 6 mostrar uma diferença na probabilidade de inadequação no consumo deste nutriente, esse consumo se mostrou independente das covariáveis.

A tabela 13 mostra o ajuste do modelo proposto pelo método NCI do consumo de vitamina B12 para os idosos do sexo masculino e feminino.

Tabela 13. Ajuste do modelo com erro de medida proposto pelo método NCI para o consumo de vitamina B12 estratificado por sexo. Avaré, 2009.

Vitamina B12				
Sexo	Covariáveis	Estimativas	Erro Padrão	p
Masculino	Intercepto	9,79	3,64	0,01
	Idade	-0,07	0,03	0,03
	Índice de massa corpórea	-0,10	0,06	0.14
	Eutrofíco	-0,04	0,61	0,95
	Sobrepeso	-0,42	0,82	0,61
	Uso de prótese	0,25	0,36	0,51
	Faz uso de suplemento	-0,48	0,89	0,59
	Possui problemas intestinais	-0,52	0,25	0,05
	Tabagismo	0,09	0,44	0,84
	Faz atividade física	0,50	0,41	0,23
	Transformação Box-Cox(λ)	0,26	0,10	0,02
	Variância interpessoal	-0,06	0,13	0,62
	Variância intrapessoal	-0,89	0,43	0,05
Feminino	Intercepto	-0,09	2,14	0,97
	Idade	-0,04	0,02	0,13
	Índice de massa corpórea	0,02	0,05	0,71
	Eutrofíco	0,19	0,49	0,71
	Sobrepeso	-0,32	0,83	0,70
	Uso de prótese	0,79	0,77	0,31
	Faz uso de suplemento	-0,70	0,30	0,03
	Possui problemas intestinais	0,21	0,14	0,15
	Tabagismo	1,64	0,86	0,06
	Faz atividade física	-0,32	0,28	0,27
	Transformação Box-Cox(λ)	0,01	-	-
	Variância interpessoal	-0,02	0,09	0,81
	Variância intrapessoal	-0,99	0,46	0,04

A tabela 13, mostra que houve influência significativa da idade para o consumo de vitamina B12 para o sexo masculino. Já o uso de suplementos alimentares e a variabilidade intrapessoal tiveram influência significativa no consumo de vitamina B12 para o sexo feminino. Isto influencia na probabilidade de inadequação do consumo desse nutriente, conforme obtido na tabela 6.

A tabela 14 mostra o ajuste do modelo proposto pelo método NCI do consumo de vitamina C para os idosos do sexo masculino e feminino.

Tabela 14. Ajuste do modelo com erro de medida proposto pelo método NCI para o consumo de vitamina C estratificado por sexo. Avaré, 2009.

Vitamina C				
Sexo	Covariáveis	Estimativas	Erro Padrão	p
Masculino	Intercepto	4,93	4,03	0,23
	Idade	0,04	0,03	0,29
	Índice de massa corpórea	0,01	0,07	0,84
	Eutrofíco	-0,90	0,68	0,20
	Sobrepeso	-1,21	0,91	0,19
	Uso de prótese	-0,48	0,41	0,25
	Faz uso de suplemento	0,45	1,00	0,65
	Possui problemas intestinais	-0,06	0,28	0,82
	Tabagismo	-0,64	0,49	0,21
	Faz atividade física	-1,42	0,46	0,01
	Transformação Box-Cox(λ)	0,01	-	-
	Variância interpessoal	0,03	0,10	0,75
	Variância intrapessoal	-0,76	0,41	0,08
Feminino	Intercepto	0,47	4,05	0,91
	Idade	-0,06	0,05	0,23
	Índice de massa corpórea	0,02	0,11	0,84
	Eutrofíco	0,23	0,92	0,80
	Sobrepeso	-1,31	1,63	0,43
	Uso de prótese	2,12	1,51	0,17
	Faz uso de suplemento	0,30	0,58	0,61
	Possui problemas intestinais	0,27	0,27	0,33
	Tabagismo	2,61	1,68	0,13
	Faz atividade física	-0,16	0,53	0,76
	Transformação Box-Cox(λ)	0,04	0,05	0,48
	Variância interpessoal	0,35	0,26	0,18
	Variância intrapessoal	-0,03	0,30	0,93

Pela tabela 14, apenas a atividade física teve influência significativa no consumo de vitamina C para os idosos do sexo masculino, mas este fato não têm uma influência na probabilidade de inadequação (tabela 6). Já para o sexo feminino, apesar de não ter sido verificada a influência das covariáveis e da variabilidade intrapessoal, nota-se que a probabilidade de inadequação pelo método ISU é quase a metade que a probabilidade de inadequação dada pelo método NCI.

Observando as estatísticas descritivas da tabela 5, nota-se que o consumo para o sexo feminino apresenta uma variabilidade maior que para o sexo masculino, já que para o sexo feminino, existe um consumo excessivo desse nutriente. Talvez esse fato possa acarretar algum problema na obtenção de probabilidade de inadequação pelo método NCI, uma vez que a distribuição de consumo é obtida por reamostragem e a chance dessa observação discrepante ser escolhida pode ser muito baixa, o que pode interferir na distribuição de consumo e consequentemente, na probabilidade de inadequação, como pode ser visto na tabela 6.

A tabela 15 mostra o ajuste do modelo proposto pelo método NCI do consumo de vitamina E para os idosos do sexo masculino e feminino.

Tabela 15. Ajuste do modelo com erro de medida proposto pelo método NCI para o consumo de vitamina E estratificado por sexo. Avaré, 2009.

Vitamina E				
Sexo	Covariáveis	Estimativas	Erro Padrão	p
Masculino	Intercepto	4,50	1,83	0.02
	Idade	-0,02	0,02	0,31
	Índice de massa corpórea	0,00	0,03	0,96
	Eutrofíco	0,69	0,31	0,03
	Sobrepeso	0,28	0,41	0,50
	Uso de prótese	0,27	0,18	0,15
	Faz uso de suplemento	-0,52	0,45	0,26
	Possui problemas intestinais	-0,11	0,13	0,40
	Tabagismo	-0,43	0,22	0,07
	Faz atividade física	0,09	0,21	0,67
	Transformação Box-Cox(λ)	0,01	-	-
	Variância interpessoal	-0,88	0,10	0,00
	Variância intrapessoal	-1,39	0,29	0,00
Feminino	Intercepto	1,70	1,17	0,15
	Idade	-0,01	0,01	0,31
	Índice de massa corpórea	-0,06	0,03	0,07
	Eutrofíco	0,19	0,27	0,49
	Sobrepeso	0,58	0,46	0,22
	Uso de prótese	0,70	0,44	0,13
	Faz uso de suplemento	0,08	0,17	0,63
	Possui problemas intestinais	-0,03	0,08	0,70
	Tabagismo	0,95	0,51	0,07
	Faz atividade física	0,02	0,15	0,91
	Transformação Box-Cox(λ)	0,07	0,13	0,60
	Variância interpessoal	-0,63	0,26	0,02
	Variância intrapessoal	-1,62	0,54	0,01

Pela tabela 15, pode-se observar que a eutrofia influenciou no consumo de vitamina E para o sexo masculino e as variabilidades intra e interpessoal tiveram influência significativa para ambos os sexos, indicando que não existiu um padrão de consumo para esse nutriente.

5 CONCLUSÕES

Com os resultados obtidos neste estudo concluiu-se que:

i) a maioria dos nutrientes analisados para idosos apresenta uma alta probabilidade de inadequação de consumo, tanto pelo método ISU quanto pelo método NCI;

ii) a estimação da distribuição de consumo de nutrientes não mostrou ser uma tarefa simples de se obter tanto pelo método ISU quanto pelo método NCI, já que, para alguns nutrientes, as probabilidades de inadequação de consumo podem ser muito diferentes mesmo não havendo influência das covariáveis, adicionado a fatores referentes à veracidade da informação, problemas de digitação, falta de memória dos informantes, dentre outras informações que podem afetar de maneira direta ou indireta o consumo;

iii) apesar do tamanho amostral considerado no presente trabalho ser considerado pequeno, uma estratégia para melhorar a estimação do consumo habitual para alguns nutrientes cujo consumo pode ser considerado esporádico ou aqueles que apresentam alta variabilidade no consumo, seria, a partir das variabilidades inter e intrapessoais, obter o número mínimo de recordatórios necessário para que a probabilidade de inadequação seja estimada com precisão e confibilidade;

iv) apesar do modelo utilizado em ambos os métodos ser com erro de medida, esperaríamos que a distribuição estimada do consumo com as devidas probabilidades de inadequação fossem próximas quando não houvesse influência das covariáveis. Isto implica que outros estudos devem ser conduzidos talvez com outras covariáveis que podem ter uma influência significativa no consumo como questões referentes à dependência e capacidade funcional, atividades de vida diária, dentre

outras.

REFERÊNCIAS BIBLIOGRÁFICAS

BEATON, G. H.; MILNER, J.; COREY, P. Sources of variance in 24-hour dietary recall data: implications for nutrition study design and interpretation. **American Jornal Clinical Nutrition**, v.32, n.12, p.2546–2559, 1972.

CASELLA, G.; BERGER, R. L. Regression Models. In: **Statistical Inference**. New York: Duxbury Press, 2001. p.577–591.

COUNCIL, N. R. **Nutrient adequacy:assessment using food consumption surveys**. Washington (DC): National Academy Press, 1986. 146p.

DODD, K. W. National Cancer Institute. **Personal communication**, 2004.

DODD, K. W.; GUENTHER, P. M. Statistical methods for estimating usual intake of nutrients and foods: a review of the theory. **Journal of the American Dietetic Association**, v.106, n.10, p.1640–1650, 2006, doi:10.1016/j.jada.2006.07.011.

DRI. Dietary Reference Intakes: Applications in dietary assessment. **Institute Medicine, Food and Nutrition Board: Avaiable at: http://www.nap.edu/books/0309071836/html.**, 2006.

FISBERG, R.; SLATER, B.; MARCHIONI, D. M. L.; MARTIN, L. A. Métodos de inquéritos alimentares. In: **Inquéritos Alimentares - Métodos e bases científicas**. Brasil: Manole, 2005. p.1–32.

GIBSON, R. S. Principles of nutritional assessment. In: **Food consumption of individuals**. Oxford: Oxford University Press, 1990. p.37–54.

HEGSTED, D. M. The classic approach - The USDA nationwide food consumption survey. **The American Jornal Clinical Nutrition**, v.35, n.5, p.1302–1305, 1972a.

HEGSTED, D. M. Problems in the use and interpretation of the recommeded daily allowances. **Ecology of Food and Nutrition**, v.1, p.255–265, 1972b.

NUSSER, S. M.; DODD, K. W.; CARRIQUIRY, A. L.; FULLER, W. A. A semi-parametric transformation approch to estimating usual nutrient intake distributions. **J Am Stat Assoc**, v.91, p.1440–1449, 1996a.

NUSSER, S. M.; FULLER, W. A.; GUENTHER, P. M. Estimation of usual dietary intake distributions: Adjusting for measurement error and nonnormality in 24-hour food intake data. In: TREWIN, D. (Ed.). **Survey Measurement and Process Quality**. New York: NY: Wiley, 1996b. p.689–709.

SLATER, B.; MARCHIONI, S. T.; FISBERG, R. M. Validação de questionários de frequência alimentar - QFA - considerações metodológicas. **Revista Brasileira de Epidemiologia**, v.6, n.3, p.2000–2008, 2003.

SOLARI, M. E. The "maximum likelihood solution"to the problem of estimating a linear functional relationship. **J R Statist Soc**, v.31, p.372–375, 1969.

TOOZE, J. A.; MIDTHUNE, D.; DODD, K. W.; FREEDMAN, L. S.; KREBS-SMITH, S. M.; SUBAR, A. F.; GUENTHER, P. M.; CARROLL, R. J.; KIPNIS, V. A new statistical method for estimating the usual intake of episodically consumed foods with application to their distribution. **Jornal American Dietetics Association**, v.106, n.10, p.1575–1587, 2006a.

TOOZE, J. A.; MIDTHUNE, D.; DODD, K. W.; FREEDMAN, L. S.; KREBS-SMITH, S. M.; SUBAR, A. F.; GUENTHER, P. M.; CARROLL, R. J.; KIPNIS, V. Statistical methods for Estimating Usual Intake of Nutrients and Foods: A

Review of the Theory. **Jornal American Dietetic Association**, v.106, n.10, p.1640–1650, 2006b.

WILLETT, W. C. **Nutritional Epidemiology**. New York: Oxford University Press, 1998. 212p.