

# Use of Asymmetric Models to Adjust the Vitamin Intake Distribution Data for Older People

José Eduardo Corrente<sup>1</sup>, Giovana Fumes<sup>2</sup>

<sup>1</sup>Biostatistics Department, Bioscience Institute, University of Sao Paulo State, Botucatu, Brazil

<sup>2</sup>Department of Exact Sciences, "Luiz de Queiroz" College of Agriculture, Botucatu, Brazil

Email: jecorren@ibb.unesp.br, gifumesbtu@gmail.com

Received 23 May 2016; accepted 18 June 2016; published 22 June 2016

Copyright © 2016 by authors and Scientific Research Publishing Inc.

This work is licensed under the Creative Commons Attribution International License (CC BY).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

---

## Abstract

One of the main interests in the nutrition field is to estimate the distribution of usual nutrient intake. Data from vitamin intake generally present high asymmetry mainly to the presence of outliers. This can occur due to the variability of the diet and, in this case, robust estimation to get the distribution of the data can be required. Then, the aim of paper is to propose an alternative approach for estimating usual intake through asymmetric distributions with random effects applied to data set 10 vitamins obtained from a dietetic survey for 368 older people from Botucatu city, São Paulo, Brazil. In general, these asymmetric distributions include parameters related to mean, median, dispersion measures and such parameters provide good estimates for the intake distribution. In order to make some comparisons, a model fitted by National Cancer Institute (NCI) method with only for amount of nutrient intake was established using Akaike Information Criteria (AIC). NCI method is based on a Box-Cox transformation coupled with normal distribution but in case of asymmetric data, this transformation can be not useful. It was observed that, in the presence of outliers, the asymmetric models provided a better fit than the NCI method in the major of the cases. Then, these models can be an alternative method to estimate the distribution of nutrient intake mainly because a transformation for the data is no necessary and all the information can be obtained directly from the parameters.

## Keywords

Asymmetric Models, Older People, Vitamin Intake

---

## 1. Introduction

One of the main interests of the researchers in the nutrition area is to estimate the distribution of usual nutrient

intakes of a population group.

The statistical modelling to measure the distribution of usual nutrient intake of a population presents a challenge, since different individuals tend to have different dietary habits from each other (between-person variability), and the individual himself does not necessarily have a constant consumption (within-person variability) [1].

In the literature, it is common to use a Box & Cox transformation [2] and analyze consumption data under normal assumption. Methods such as Iowa State University (ISU) and Best Power (BP) follow this background [3] as they do not allow the inclusion of covariates. Such methodologies take in account the normality of the data before or after a transformation and generally differ on back transformation to be applied to get an original scale data, as well as the form of attenuating weight factor considered in such a context. Another model, called the National Cancer Institute (NCI) method [4], allows the inclusion of covariates and takes into account the probability of an episodic intake. However, for data with high asymmetry and/or presence of extreme outliers, the NCI-based modeling method doesn't seem to be very appropriate.

Normal distribution is one of the most used distributions in the analysis of continuous data due to its properties, especially in the context of linear models. However, the presence of outliers affects relevantly the inference based on the normal distribution, encouraging the development of robust procedures, which are defined as less sensitive to deviations from the assumptions which they are based on [5]. Besides, the use of asymmetric models can be an alternative form to model consumption data without the need of a transformation in order to get normal distribution [6].

Thus, the aim of this paper is to obtain an estimate for the usual consumption of vitamins for older people through the application of asymmetric models, and some of them may contemplate a robust estimation procedure (in case of strong presence of aberrant points). Furthermore, a comparison with the NCI method was made considering only the amount consumed [7].

## 2. Methodology

The analyzed data came from a representative sample of 368 individuals belonging to an epidemiological study that aimed to assess the adequacy of nutrient intake for older people from Botucatu city, São Paulo, Brazil.

The sample was collected in the year of 2011 through the application of the 24-hour recall (R24h). The study included elders who do not have cognitive deficits and agreed to participate into the research. In order to estimate the usual intake of the subject, more than one R24h is needed, due to the variability in consumption on different days of the week. Thus, up to three R24h were obtained from each subject on nonconsecutive days of the week, one of them being obtained necessarily on the weekend.

The data collected by the recalls were transformed into micronutrient consumption data using the NDSR program (Nutrition Data System for Research).

Nutrient intakes are different for men and women. Thus, the analyses were performed by gender.

The data were also collected considering age, marital status, education and morbidities as well as anthropometry measures and information about the Activities of Daily Living (ADL) and the Instrumental Activities of Daily Living (IADL).

A descriptive analysis of ten vitamin consumptions were presented as: mean, standard deviation (SD), median and coefficient of variation (CV\*) based on the median of the data distribution [8].

The data modeling was performed by means of a random intercept model, which characterizes different measurements on the same subject. The idea of the used model is that the between-person variability is absorbed by a random effect, and the within-person variability is absorbed by the nature of the chosen distribution for the response variable that is similar to the NCI method [7].

Asymmetric distributions were proposed for the response variable and it was used a normal distribution with zero mean and variance  $\kappa^2$  for the random effect. In order to select such asymmetric distributions, the *fitdist* and *histdist* functions from *Gamlss* (Generalized Additive Models for Location, Scale and Shape) routine at R software, v.3.0.1, were used [9]. The penalized maximum likelihood method was used to estimate the parameters of the asymmetric models. Furthermore, the estimate processes were made by RS and CG interactive algorithm [10]. The variance of the random effect was estimated by *gamlss.mx* using EM algorithm. The fitted asymmetric models were made using R software, v.3.0.1 [11].

The NCI method for amount-only model presents in SAS version 9.3 software implemented by macro MIXTRAN was used for comparison with the proposed models, since it has a similar structure with the asym-

metric models in this study. The average usual intake consumption estimated by the NCI method using the macro DISTRIB, is also shown for the comparison purpose with the parameters estimated by asymmetric models.

### 3. Results

The sample of 368 older people presented mean age of  $71.20 \pm 7.11$  years old for men and  $71.79 \pm 7.21$  years old for women. **Table 1** presents the description of demographic and morbidity variables of the elders.

**Table 2** presents the descriptive measures for the vitamin intakes of the older people according to gender.

**Table 2** shows that the standard deviations (SD) were too high for most vitamins indicating a high dispersion around the mean. Moreover, average and median values for most of the vitamins are not close, indicating an asymmetry in the data. Thus, the reported descriptive measures suggest a proposal for statistical modeling that includes an asymmetry and robust estimation.

Gama distribution, generalized Gamma, Weibull, Lognormal and Box-Cox t were adjusted to the response variable. The adjusted distributions may have an asymmetric shape and leptokurtic, depending on the values assigned to their parameters. These distributions were selected through *fitdist* and *histdist* functions, which selected the best distribution from the raw data. On the features of these distributions, the generalized gamma distribution is a specific parameterization of the gamma distribution proposed by Lopatzidis and Green, in which it has an additional parameter  $\nu$  [8] [9]. The parameters  $\mu$  and  $\sigma$  to the aforementioned distribution are associated with the mean and variance respectively. In the Box-Cox t (BCT) distribution the  $\mu$ ,  $\sigma$ ,  $\nu$  and  $\tau$  parameters can be interpreted as scale (related to median), relative dispersion (associated with the coefficient of variation), asymmetry (processing power for symmetry) and kurtosis (degree of freedom), respectively. The lognormal distribution is a particular case of Box-Cox Cole Green (BCCG) distribution when the parameter  $\nu = 0$ . Being the BCCG distribution a particular case of BCT, when the parameter for the number of degrees of freedom tends to

**Table 1.** Distribution of the elders according to demographic and morbidities variables by gender (n = 368) (Botucatu, 2011).

Variable	Category	Men (%) (n = 136)	Women (%) (n = 232)	p-value
Marital status	Not married	18.1	54.0	0.0001
	Married	81.9	46.0	
Education	Until primary school	66.1	81.6	0.0003
	Above primary school	33.9	18.4	
Hypertension	Yes	51.7	58.9	0.2644
	No	48.3	41.1	
Diabetes	Yes	29.1	27.8	0.4447
	No	70.9	72.2	
Cholesterol	Yes	12.5	16.8	0.2665
	No	87.5	83.2	
Cardiovascular disease	Yes	6.9	8.1	0.6722
	No	91.9	93.1	
BMI <sup>1</sup>	Underweight	20.6	17.7	0.4487
	Normal	32.3	28.4	
	Overweight	47.1	53.9	
ADL <sup>2</sup>	Independent	86.7	80.3	0.1194
	Dependent	6.8	10.3	
IADL <sup>3</sup>	Independent	70.0	59.1	0.5285
	Dependent	23.5	31.5	

<sup>1</sup>Body mass index; <sup>2</sup>Activities of daily living; <sup>3</sup>Instrumental activities of daily living.

**Table 2.** Descriptive measures for the vitamin intakes for the older people (n = 368) (Botucatu, 2011).

Vitamins	Men (n = 136)				Women (n = 232)				p-value
	Mean	SD	Median	CV*	Mean	SD	Median	CV*	
A (mcg)	1249.71	3163.41	713.22	0.91	1135.33	1712.59	756.86	0.97	0.0001
D (mcg)	4.33	3.42	3.64	0.71	3.94	2.83	3.40	0.70	0.0001
E (mg)	7.55	12.61	6.13	0.4	6.13	3.57	5.48	0.48	0.0001
K (mcg)	241.11	1656.86	103.37	0.96	196.00	1145.12	101.77	0.85	0.0001
C (mg)	209.41	856.62	54.72	1.53	104.98	269.25	54.08	1.19	0.0001
B1 (mg)	1.78	1.23	1.53	0.40	1.48	0.74	1.34	0.44	0.0001
B2 (mg)	1.69	0.91	1.55	0.48	1.52	0.72	1.41	0.41	0.0001
B3 (mg)	24.23	18.22	20.60	0.53	19.83	11.37	17.50	0.55	0.0001
B6 (mg)	2.08	2.09	1.77	0.49	1.67	0.82	1.52	0.52	0.0620
B12 (mcg)	6.25	10.73	3.47	0.77	7.54	18.20	2.93	0.79	0.0001

\* p-value from mann-whitney test.

infinity, it results in a truncated normal distribution [12]. For the random effect, a normal distribution was proposed with zero mean and variance  $\kappa^2$ .

**Table 3** presents fitted models to a vitamin intake data for older people and the estimate parameters.

In **Table 3**, it was noted that the parameter estimate related to the median of the data for Box-Cox t distribution lies very close to the observed median from the data, and in some cases slightly lower (see **Table 2**), as it was expected for this parameter to be composed of a random intercept model. For other models in which that  $\mu$  is a parameter related to the mean of the intake distribution, the estimate of this parameter from the adjusted model was also very close to the observed average (see **Table 2**). For instance, according to the gamma model adjusted for consumption of vitamin B1 in men, the average estimated from the model was 1.66 mg, while the average found in the database was 1.78 mg. It is also observed that in the Box-Cox t distribution, which has a robustness in the estimation process, the parameter for the number of degrees of freedom presents small estimates suggesting heavier tails to accommodate outliers. It is a very important fact to model intake data, which are very different. For example, according to the data from Vitamin A in men, the lowest consumption was 1.86 mcg, while the largest one was 41,372.02 mcg.

The Akaike criterion value was used to compare the NCI method for vitamin intakes and the asymmetric distributions. By this criterion, it was observed that there was no difference between the models, although most of the asymmetrical models had a lower value. It was not considered the adjustment of energy for the evaluated nutrients, since the Akaike values were not very different from those without adjustments. The results are shown in **Table 4**.

It is important to notice that the NCI method presented limitations for the adjustment of some vitamins as it uses a Box-Cox transformation type. When this transformation is not found, the result in MIXTRAN macro uses a logarithmic transformation to solve the problem, but this does not guarantee that this assumption can lead data to normal distribution. Asymmetric models also showed problems in the estimation of some models by the *gamlss* routine. When this problem happened, alternative distributions were selected through *histdist* and *fitdist*-functions.

## 4. Discussion

Several methods to estimate the usual intake have been proposed in literature such as National Cancer Institute (NCI—considered as a standard method in this work), Multiple Source Method (MSM), Iowa State University (ISU) and Statistical Program for Age-adjusted Dietary Assessment (SPADE) [13] [14]. The authors carried out a simulated study with two 24-hour recalls with unequal number of subjects and they concluded that care should be taken if there is a high between-person variability or a high asymmetry.

The use of asymmetric distributions was already proposed but without taking in to account the between and

**Table 3.** Fitted asymmetric models to a vitamin intakes distribution for older people according to gender (n = 368). Botucatu, 2011.

Gender	Vitamins	Fitted model	Parameter estimates
Male	A (mcg)	Box-Cox t	$\hat{\mu} = 687.30, \hat{\sigma} = 0.81, \hat{\nu} = 0.10, \hat{\tau} = 4.53, \hat{\kappa} = 0.60$
	D (mcg)	Gama	$\hat{\mu} = 3.71, \hat{\sigma} = 0.53, \hat{\kappa} = 0.46$
	E (mg)	Box-Cox t	$\hat{\mu} = 5.97, \hat{\sigma} = 0.34, \hat{\nu} = -0.05, \hat{\tau} = 3.71, \hat{\kappa} = 0.25$
	K (mcg)	Box-Cox t	$\hat{\mu} = 100.69, \hat{\sigma} = 0.50, \hat{\nu} = 0.00, \hat{\tau} = 3.77, \hat{\kappa} = 0.52$
	C (mg)	Box-Cox t	$\hat{\mu} = 53.45, \hat{\sigma} = 1.19, \hat{\nu} = 0.29, \hat{\tau} = 2.24, \hat{\kappa} = 0.71$
	B1 (mg)	Gama	$\hat{\mu} = 1.66, \hat{\sigma} = 0.33, \hat{\kappa} = 0.30$
	B2 (mg)	Lognormal	$\hat{\mu} = 0.42, \hat{\sigma} = 0.31, \hat{\kappa} = 0.27$
	B3 (mg)	Box-Cox t	$\hat{\mu} = 20.87, \hat{\sigma} = 0.45, \hat{\nu} = 0.10, \hat{\tau} = 5.40, \hat{\kappa} = 0.18$
	B6 (mg)	Generalized Gama	$\hat{\mu} = 1.56, \hat{\sigma} = 0.38, \hat{\nu} = -1.47, \hat{\kappa} = 0.28$
	B12 (mcg)	Generalized Gama	$\hat{\mu} = 2.84, \hat{\sigma} = 0.67, \hat{\nu} = -0.98, \hat{\kappa} = 0.44$
Female	A (mcg)	Box-Cox t	$\hat{\mu} = 710.70, \hat{\sigma} = 0.85, \hat{\nu} = 0.10, \hat{\tau} = 6.48, \hat{\kappa} = 0.53$
	D (mcg)	Weibull	$\hat{\mu} = 3.82, \hat{\sigma} = 2.21, \hat{\kappa} = 0.44$
	E (mg)	Lognormal	$\hat{\mu} = 1.70, \hat{\sigma} = 0.40, \hat{\kappa} = 0.27$
	K (mcg)	Box-Cox t	$\hat{\mu} = 96.32, \hat{\sigma} = 0.78, \hat{\nu} = -0.00, \hat{\tau} = 10.18, \hat{\kappa} = 0.43$
	C (mg)	Box-Cox t	$\hat{\mu} = 54.57, \hat{\sigma} = 0.95, \hat{\nu} = 0.18, \hat{\tau} = 4.64, \hat{\kappa} = 0.61$
	B1 (mg)	Generalized Gama	$\hat{\mu} = 1.27, \hat{\sigma} = 0.36, \hat{\nu} = -0.71, \hat{\kappa} = 0.19$
	B2 (mg)	Lognormal	$\hat{\mu} = 0.32, \hat{\sigma} = 0.30, \hat{\kappa} = 0.25$
	B3 (mg)	Lognormal	$\hat{\mu} = 2.84, \hat{\sigma} = 0.45, \hat{\kappa} = 0.22$
	B6 (mg)	Gama	$\hat{\mu} = 1.60, \hat{\sigma} = 0.39, \hat{\kappa} = 0.26$
	B12 (mcg)	Gama	$\hat{\mu} = 4.43, \hat{\sigma} = 0.78, \hat{\kappa} = 0.79$

**Table 4.** Akaike criteria for the NCI and asymmetric models for the vitamin intakes for older people. Botucatu, 2011.

Gender	Male		Female	
	Vitamins	NCI	NCI	Asymmetric models
	A (mcg)	5641.90	<b>5614.30</b>	9426.02
	D (mcg)	1637.30	<b>1593.65</b>	2630.90
	E (mg)	1861.90	<b>1812.65</b>	2879.90
	K (mcg)	4244.20	<b>4167.70</b>	6951.60
	C (mg)	4011.20	<b>3974.29</b>	6535.30
	B1 (mg)	783.70	<b>743.53</b>	1062.40
	B2 (mg)	719.70	<b>660.30</b>	1042.20
	B3 (mg)	2747.50	<b>2730.01</b>	4330.40
	B6 (mg)	967.00	<b>913.07</b>	1310.40
	B12 (mcg)	1871.00	<b>1823.47</b>	3229.78

within-person variability [6]. In this case, some asymmetric distributions were proposed and, even without such variability, the results were good mainly to estimate the prevalence of inadequacy for the considered nutrients.

Another class of functions, called Box-Cox symmetric class [12] has been developed in order to model such intake data with appropriate routines implemented by the author. Again, using the Akaike criteria, these proposed models presented, in general, lower values than the NCI method, indicating an improvement in the adjustment of the usual intake distribution of the vitamins. The energy adjust did not modify the results as well as the adjust for possible confounder variables.

In both methods (NCI and asymmetric models), it was observed that, due to the presence of high asymmetry and outliers, the variability of the random effect is high with or without energy adjustment. This means that the between-person variability exists and shows that there is not a nutrient pattern of consumption.

Therefore, this work shows that the adjustment of asymmetric models for nutritional intake data are effective and have, as an advantage, the direct calculation of the mean and median intake using fitted distribution without the need of a transformation, as made in the NCI method. Another advantage is the practicality of application, due to the tools presented in the `gamlss` routine. However, for future studies, it is intended to implement such asymmetric distributions to improve the issue of parameter estimation, and eliminate the problems related to the lack of convergence found in the used routine. Moreover, with such models implemented, it is possible to estimate the prevalence of inadequacy of nutrient intakes based on a cutoff point fixed a priori.

## 5. Conclusion

In this work, the use of asymmetric distribution to fit the nutrient intake distribution was proposed based on a random effect model in case of outliers and models with robust estimation procedure. An advantage of this new approach is that no data transformation is needed and the results can be interpreted directly from the estimated parameters. In addition, the inclusion of outliers is possible by means of robust procedures providing more plausible estimates.

## Acknowledgements

The authors would like to thank the São Paulo Research Foundation (FAPESP-Process no. 2008/10261-8) and the National Council for Scientific and Technological Development (CNPq-Process no. 301197/2011-3) for the financial support for this research.

## References

- [1] Borrelli, R., Simonetti, M.S. and Fidanza, F. (1992) Inter- and Intra-Individual Variability in Food Intake of Elderly People in Perugia (Italy). *British Journal of Nutrition*, **68**, 3-10. <http://dx.doi.org/10.1079/BJN19920061>
- [2] Box, G.E.P. and Cox, D.R. (1964) An Analysis of Transformations. *Journal of the Royal Statistical Society B*, **26**, 211-252.
- [3] Dodd, K.W., Guenther, P., Freedman, L., Subar, A., Kipnis, V., Midthune, D., Tooze, J. and Smith, S. (2006) Statistical Methods for Estimating Usual Intake of Nutrients and Foods: A Review of the Theory. *Journal of the American Dietetic Association*, **106**, 1640-1650. <http://dx.doi.org/10.1016/j.jada.2006.07.011>
- [4] Tooze, J.A., Midthune, D., Dood, K.W., Freedman, L.S., Krebs-Smith, S.M., Subar, A.F., Guenther, P.M., Carroll, R.J. and Kipnis, V. (2006) A New Statistical Method for Estimating the Usual Intake of Episodically Consumed Foods with Application to Their Distribution. *Journal of the American Dietetic Association*, **106**, 1575-1587. <http://dx.doi.org/10.1016/j.jada.2006.07.003>
- [5] Hampel, F.R., Ronchetti, E.M., Rousseeuw, P.J. and Stahel, W.A. (1986) *Robust Statistics: The Approach Based on Influence Functions*. John Wiley & Sons, New York.
- [6] Corrente, J.E., Marimoto, J., Marchioni, D.M.L. and Fisberg, R.M. (2011) Alternative Distributions to Estimate Usual Intake of Nutrients for Groups. *Journal of Life Sciences*, **5**, 569-574.
- [7] Tooze, J.A., Kipnis, V., Buckman, D.W., Carroll, R., Freedman, L., Guenther, P., Krebs-Smith, S., Subar, A. and Dodd, K. (2010) A Mixed-Effects Model Approach for Estimating the Distribution of Usual Intake of Nutrients: The NCI Method. *Statistics in Medicine*, **29**, 2857-2868. <http://dx.doi.org/10.1002/sim.4063>
- [8] Rigby, R.A. and Stasinopoulos, D.M. (2006) Using the Box-Cox t Distribution in GAMLSS to Model Skewness and Kurtosis. *Statistical Modelling*, **6**, 209-229. <http://dx.doi.org/10.1191/1471082X06st122oa>
- [9] Rigby, R.A. and Stasinopoulos, D.M. (2006) *Statistical Modelling Using GAMLSS in R*. London. <http://www.gamlss.org>

- 
- [10] Rigby, R.A. and Stasinopoulos, D.M. (2005) Generalized Additive Models for Location, Scale and Shape. *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, **54**, 507-554. <http://dx.doi.org/10.1111/j.1467-9876.2005.00510.x>
- [11] Stasinopoulos, D.M., Rigby, R.A. and Akantziliotou, C. (2008) Instructions on How to Use the GAMLSS Package in R. London. <http://www.gamlss.org>
- [12] Fumes, G. (2015) Asymmetric Box-Cox Models and Application to Nutritional Data. PhD Thesis, University of São Paulo, São Paulo.
- [13] Souverein, O.W., Dekkers, A.L., Geelen, A., Hwubrock, J., de Vries, J.H., Ockré, M.C., Harttig, U., Boeing, H. and Veer, P.V. (2011) Comparing Four Methods to Estimate Usual Intake Distribution. *European Journal of Clinical Nutrition*, **65**, S92-S101. <http://dx.doi.org/10.1038/ejcn.2011.93>
- [14] Waijers, P.M.C.M., Dekkers, A.L.M., Boer, J.M.A., Boshuizen, H.C. and Van Rossum, C.T.M. (2006) The Potential of AGE MODE, an Age-Dependent Model, to Estimate Usual Intakes and Prevalences of Inadequate Intakes in a Population. *The Journal of Nutrition*, **136**, 2916-2920.