Estudo de simulação para comparar métodos de estimação da distribuição do consumo alimentar usual: ISU, NCI, MSM e SPADE.

G H C Laureano[1]
V B L Torman[2]
S P Crispim[3]
A L M Dekkers[4]
S A Camey[5]

Resumo: Muitos métodos estão disponíveis para estimar as distribuições de consumo alimentar usual e, por isso, há uma necessidade de estudos de simulação para compará-los. Os métodos ISU (*Iowa State University*), NCI (*National Cancer Institute*), MSM (*Multiple Source Method*), e SPADE (*Statistical Program to Assess Dietary Exposure*) foram comparados anteriormente em um estudo realizado por SOUVEREIN et al., 2011, mas alguns resultados não foram conclusivos devido ao pequeno número de replicações utilizadas na simulação. Buscando superar esta limitação, o presente estudo utilizou 1.000 replicações na simulação para doze cenários diferentes, com foco em amostras pequenas e grandes razões de variâncias, para comparar a precisão e acurácia das estimativas geradas pelos métodos acima mencionados. O vício e o vício relativo foram utilizados como medidas de acurácia, enquanto que o erro quadrático médio foi utilizado como a medida da precisão. Para pequenas amostras, os métodos ISU, MSM e SPADE obtiveram estimativas mais acuradas e precisas do que o método NCI, particularmente para os percentis 5 e 95.

Palavras-chave: *Distribuição de consumo alimentar usual, ISU, NCI, MSM, SPADE.*

Introdução

A mensuração do consumo alimentar usual é assunto vigente na área da saúde, uma vez que várias doenças são influenciadas, ou até mesmo causadas, pelos hábitos alimentares do indivíduo (DONG et al., 2011). Uma das formas de estudar o consumo alimentar de uma população é por meio da estimação da distribuição do consumo alimentar usual. O estudo dessas distribuições traz um melhor entendimento a respeito do consumo alimentar em populações, como também, pode ajudar na identificação de possíveis grupos de risco para o desenvolvimento de doenças.

Foram propostas diversas metodologias de estimação dessa distribuição (GAY, 200AD; SLOB, 1993, 2006; WALLACE; DUAN; ZIEGENFUS, 1994; BUCK; HAMMERSTROM; RYAN, 1995; NUSSER et al., 1996; GUENTHER; KOTT; CARRIQUIRY, 1997; HOFFMANN et al., 2002; TOOZE; GRUNWALD; JONES, 2002; TOOZE et al., 2006, 2010; WAIJERS et al., 2006; KIPNIS et al., 2009; HAUBROCK et al., 2011; DEKKERS; SLOB, 2012; NUSSER; FULLER; GUENTHER, 2012), e que alguns autores realizaram estudos onde

[1]Programa de pós graduação em Epidemiologia, Universidade Federal do Rio Grande do Sul (UFRGS). Email: greice.laureano@ufrgs.br
[2]Programa de pós graduação em Epidemiologia e Departamento de Estatística, UFRGS. Email: vanessa.leotti@ufrgs.br
[3]Departamento de Nutrição, Universidade Federal do Paraná (UFPR). Email: crispim@ufpr.br
[4]Netherlands National Institute for Public Health and the Environment. Email: Arnold.Dekkers@rivm.nl
[5]Programa de pós graduação em Epidemiologia e Departamento de Estatística, UFRGS. Email: camey@mat.ufrgs.br

comparavam métodos de estimação da distribuição (GOEDHART et al., ; NUSSER et al., 1996; GUENTHER; KOTT; CARRIQUIRY, 1997; HOFFMANN et al., 2002; DODD et al., 2006; TOOZE et al., 2010; SOUVEREIN et al., 2011).

No trabalho de SOUVEREIN et al., 2011, os métodos ISU (*Iowa State University*), NCI (*National Cancer Institute*), MSM (*Multiple Source Method*) e SPADE (*Statistical Program to Assess Dietary Exposure*) foram comparados, mas alguns resultados foram inconclusivos devido ao pequeno número de replicações na simulação. Com o objetivo de superar essa limitação, este trabalho utilizou um número maior de replicações dando enfoque aos cenários que apresentaram maior incerteza nas conclusões que foram os com razões de variâncias iguais a 4 e 9, tamanhos amostrais iguais a 150, 500 e distribuições de consumo assimétricas. Para melhor entendimento de alguns resultados foi adicionado tamanho de amostra igual a 300.

O objetivo deste trabalho é comparar os métodos ISU, NCI, MSM e SPADE em um estudo de simulação e comparar suas acurácias e suas precisões na estimação da distribuição de consumo alimentar usual.

Métodos

O consumo alimentar usual, em geral, tem distribuição assimétrica (DODD et al., 2006), o que impossibilita a utilização de métodos estatísticos baseados na distribuição Normal. Nesses casos uma das estratégias utilizadas é a transformação dos dados.

Quase todos os métodos existentes (GAY, 200AD; SLOB, 1993, 2006; WALLACE; DUAN; ZIEGENFUS, 1994; BUCK; HAMMERSTROM; RYAN, 1995; NUSSER et al., 1996; GUENTHER; KOTT; CARRIQUIRY, 1997; HOFFMANN et al., 2002; TOOZE; GRUNWALD; JONES, 2002; TOOZE et al., 2006, 2010; WAIJERS et al., 2006; KIPNIS et al., 2009; HAUBROCK et al., 2011; DEKKERS; SLOB, 2012; NUSSER; FULLER; GUENTHER, 2012) estimam as variâncias (intraindividual e interindividual) e a média do consumo da população, na escala transformada (normal). Posteriormente faz-se a transformação inversa para a escala original e então se obtém a média e os percentis da distribuição do consumo alimentar usual.

Esses métodos diferem em relação às formas de transformação, aos modelos para estimar as variâncias e a média na escala transformada, às formas de se obter a transformação inversa e aos métodos de obtenção da média e dos percentis da distribuição.

O ISU proposto pela *Iowa State University* (NUSSER et al., 1996; GUENTHER; KOTT; CARRIQUIRY, 1997; NUSSER; FULLER; GUENTHER, 2012) tem duas implementações: uma que está em SAS ("SAS | Business Analytics and Business Intelligence Software") e outra que pode ser obtida por meio de solicitação para os autores no site: http://www.side.stat.iastate.edu/pc-side.php/.

O NCI proposto pelo *National Cancer Institute* (TOOZE; GRUNWALD; JONES, 2002; TOOZE et al., 2006; KIPNIS et al., 2009; TOOZE et al., 2010), está implementado no SAS ("SAS | Business Analytics and Business Intelligence Software") e pode ser encontrado no site: http://riskfactor.cancer.gov/diet/usualintakes/.

O MSM foi proposto na Europa por uma equipe da Alemanha (HARTTIG et al., 2011; HAUBROCK et al., 2011) e está implementado na *web* no site: https://msm.dife.de/.

O SPADE (SOUVEREIN et al., 2011; DEKKERS; SLOB, 2012; DEKKERS et al., 2014) está implementado no *software* R (R CORE TEAM, 2015) e é baseado no programa AGEMODE (WAIJERS et al., 2006) onde a estimação dos consumos é modelada em função da covariável idade. A informação de idade pode ser omitida após ajustes no código do software. O SPADE é um pacote do R chamado SPADE.RIVM que pode ser obtido no endereço: www.spade.nl.

Neste estudo foram simulados 12 cenários. Na simulação foram utilizados os programas R (para a geração dos dados e para o SPADE), SAS (para o ISU e NCI) e o AutoHotkey ("AutoHotkey") (para o MSM).

Para comparação das estimativas se calculou para cada método o vício (V), o vício relativo em percentagem (VR) e o erro quadrático médio (EQM).

Para o cálculo dessas medidas foi necessário obter a média e os percentis verdadeiros e, para isso, utilizou-se a aproximação por quadratura Gaussiana (obtida pela função *f.gauss.quad* implementada no pacote SPADE.RIVM do *software* R).

Resultados

Em termos de acurácia e precisão os métodos ISU, MSM e SPADE obtiveram estimativas mais precisas e acuradas o método NCI em quase todos cenários, particularmente para os percentis 5 e 95 quando o tamanho da amostra é pequeno e a razão de variância é grande.

Discussão

Neste trabalho foi realizado um estudo de simulação que visava comparar quatro métodos utilizados para estimação da distribuição de consumo alimentar usual. Os resultados obtidos nos cenários simulados mostraram que, em termos de precisão e acurácia, o tamanho de amostra interferiu na qualidade das estimações de todos métodos.

Os resultados também mostraram que há uma estimação de pior qualidade em todos os métodos quando se trata da estimação dos percentis 5 e 95, sendo o percentil 95 o pior caso, pois obteve estimações com menor precisão.

Conclusões

Em conclusão, este estudo mostrou a importância do tamanho da amostra, que deve ser grande o suficiente para evitar problemas numéricos nos métodos utilizados.

Além disso, verificou-se também que os métodos tiveram uma performance semelhante entre si quando se utiliza tamanhos de amostra iguais a 150, 300 ou 500, apesar do método NCI ser menos preciso e acurado que os outros dois métodos.

Referências

AutoHotkey. Disponível em: <http://www.autohotkey.com/>.

BUCK, R. J.; HAMMERSTROM, K. A.; RYAN, P. B. Estimating Long-Term Exposures from Short-Term Measurements. **Journal of exposure analysis and environmental epidemiology**, v. 5, n. 3, p. 359–373, set. 1995.

DEKKERS, A. L. M.; SLOB, W. Gaussian Quadrature is an efficient method for the back-transformation in estimating the usual intake distribution when assessing dietary exposure. **Food and Chemical Toxicology**, v. 50, n. 10, p. 3853–3861, out. 2012. . Acesso em: 2 out. 2013.

DEKKERS, A. L.; VERKAIK-KLOOSTERMAN, J.; VAN ROSSUM, C. T.; OCKE, M. C. SPADE, a New Statistical Program to Estimate Habitual Dietary Intake from Multiple Food Sources and Dietary Supplements. **Journal of Nutrition**, v. 144, n. 12, p. 2083–2091, 1 dez. 2014.

DODD, K. W.; GUENTHER, P. M.; FREEDMAN, L. S.; SUBAR, A. F.; KIPNIS, V.; MIDTHUNE, D.; TOOZE, J. A.; KREBS-SMITH, S. 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, out. 2006.

DONG, J.-Y.; ZHANG, L.; HE, K.; QIN, L.-Q. Dairy Consumption and Risk of Breast Cancer: A Meta-Analysis of Prospective Cohort Studies. **Breast cancer research and treatment**, v. 127, n. 1, p. 23–31, maio 2011.

GAY, C. Estimation of population distributions of habitual nutrient intake based on a short-run weighed food diary. **British Journal of Nutrition**, p. 287–293, 200AD.

GOEDHART, P. W.; VOET, H.; SVEN, K.; DEKKERS, A. L. M.; DOOD, K. W.; BOEING, H.; KLAVEREN, J. A comparison by simulation of different methods to estimate the usual intake distribution for episodically consumed foods. **EFSA Journal**, [s.d.]

GUENTHER, P. M.; KOTT, P. S.; CARRIQUIRY, A. L. Development of an Approach for Estimating Usual Nutrient Intake Distributions at the Population Level. **The Journal of nutrition**, v. 127, n. 6, p. 1106–1112, jun. 1997.

HARTTIG, U.; HAUBROCK, J.; KNÜPPEL, S.; BOEING, H.; EFCOVAL CONSORTIUM. The MSM Program: Web-Based Statistics Package for Estimating Usual Dietary Intake Using the Multiple Source Method. **European journal of clinical nutrition**, v. 65 Suppl 1, p. S87–91, jul. 2011.

HAUBROCK, J.; NÖTHLINGS, U.; VOLATIER, J.-L.; DEKKERS, A.; OCKÉ, M.; HARTTIG, U.; ILLNER, A.-K.; KNÜPPEL, S.; ANDERSEN, L. F.; BOEING, H.; EUROPEAN FOOD CONSUMPTION VALIDATION CONSORTIUM. Estimating Usual Food Intake Distributions by Using the Multiple Source Method in the EPIC-Potsdam Calibration Study. **The Journal of nutrition**, v. 141, n. 5, p. 914–920, maio 2011.

HOFFMANN, K.; BOEING, H.; DUFOUR, A.; VOLATIER, J. L.; TELMAN, J.; VIRTANEN, M.; BECKER, W.; DE HENAUW, S.; EFCOSUM GROUP. Estimating the Distribution of Usual Dietary

Intake by Short-Term Measurements. **European journal of clinical nutrition**, v. 56 Suppl 2, p. S53–62, maio 2002.

KIPNIS, V.; MIDTHUNE, D.; BUCKMAN, D. W.; DODD, K. W.; GUENTHER, P. M.; KREBS-SMITH, S. M.; SUBAR, A. F.; TOOZE, J. A.; CARROLL, R. J.; FREEDMAN, L. S. Modeling Data with Excess Zeros and Measurement Error: Application to Evaluating Relationships between Episodically Consumed Foods and Health Outcomes. **Biometrics**, v. 65, n. 4, p. 1003–1010, dez. 2009.

NUSSER, S. M.; CARRIQUIRY, A. L.; DODD, K. W.; FULLER, W. A. A Semiparametric Transformation Approach to Estimating Usual Daily Intake Distributions. **Journal of the American Statistical Association**, v. 91, n. 436, p. 1440–1449, dez. 1996. . Acesso em: 2 out. 2013.

NUSSER, S. M.; FULLER, W. A.; GUENTHER, P. M. Estimating Usual Dietary Intake Distributions: Adjusting for Measurement Error and Nonnormality in 24-Hour Food Intake Data. In: LYBERG, L.; BIEMER, P.; COLLINS, M.; DE LEEUW, E.; DIPPO, C.; SCHWARZ, N.; TREWIN, D. (Ed.). **Survey Measurement and Process Quality**. Hoboken, NJ, USA: John Wiley & Sons, Inc., 2012. p. 689–709.

R CORE TEAM. **R: a language and environment for statistical computing**. Vienna, Austria: R Foundation for Statistical Computing, 2015.

SAS | Business Analytics and Business Intelligence Software. Disponível em: <http://www.sas.com/>.

SLOB, W. Modeling Long-Term Exposure of the Whole Population to Chemicals in Food. **Risk analysis: an official publication of the Society for Risk Analysis**, v. 13, n. 5, p. 525–530, out. 1993.

SLOB, W. Probabilistic Dietary Exposure Assessment Taking into Account Variability in Both Amount and Frequency of Consumption. **Food and chemical toxicology: an international journal published for the British Industrial Biological Research Association**, v. 44, n. 7, p. 933–951, jul. 2006.

SOUVEREIN, O. W.; DEKKERS, A. L.; GEELEN, A.; HAUBROCK, J.; DE VRIES, J. H.; OCKÉ, M. C.; HARTTIG, U.; BOEING, H.; VAN 'T VEER, P.; EFCOVAL CONSORTIUM. Comparing Four Methods to Estimate Usual Intake Distributions. **European journal of clinical nutrition**, v. 65 Suppl 1, p. S92–101, jul. 2011.

TOOZE, J. A.; GRUNWALD, G. K.; JONES, R. H. Analysis of Repeated Measures Data with Clumping at Zero. **Statistical methods in medical research**, v. 11, n. 4, p. 341–355, ago. 2002.

TOOZE, J. A.; KIPNIS, V.; BUCKMAN, D. W.; CARROLL, R. J.; FREEDMAN, L. S.; GUENTHER, P. M.; KREBS-SMITH, S. M.; SUBAR, A. F.; DODD, K. W. A mixed-effects model approach for estimating the distribution of usual intake of nutrients: The NCI method. **Statistics in Medicine**, v. 29, n. 27, p. 2857–2868, 30 nov. 2010. . Acesso em: 16 jan. 2013.

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. **Journal of the American Dietetic Association**, v. 106, n. 10, p. 1575–1587, out. 2006.

WAIJERS, P. M. C. M.; DEKKERS, A. L. M.; BOER, J. M. A.; BOSHUIZEN, H. C.; VAN ROSSUM, C. T. M. 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**, v. 136, n. 11, p. 2916–2920, nov. 2006.

WALLACE, L. A.; DUAN, N.; ZIEGENFUS, R. Can Long-Term Exposure Distributions Be Predicted from Short-Term Measurements? **Risk Analysis**, v. 14, n. 1, p. 75–85, fev. 1994. . Acesso em: 2 out. 2013.