STANDARD ERRORS OF USUAL INTAKE QUANTILES

Anthony B. An, Alicia L. Carriquiry and Wayne A. Fuller*
Anthony B. An, Department of Statistics, Iowa State University, Ames, IA 50011

ABSTRACT

Data from the 1985 Continuing Survey of Food Intakes by Individuals, conducted by the U.S. Department of Agriculture, has been used to estimate quantiles of the distribution of usual intakes for several dietary components. In this paper, we give estimators of the standard errors of the estimated quantiles. The estimation of these standard errors is not straight forward because quantiles are estimated from transformed data, and standard errors depend on the (unknown) sampling variances of a set of variance components.

KEY WORDS: Dietary usual intake, cumulation distribution function, variance components

1 INTRODUCTION

The U.S. Department of Agriculture (USDA) has conducted periodic surveys to estimate food consumption patterns of households and individuals in the United States for over 50 years. Data from these surveys have had a significant impact on the formulation of food-assistance programs, on consumer education, and on food regulatory activities.

Practitioners are interested in the distribution of usual nutrient intakes for assessing the dietary status of populations. Usual intake of a nutrient is defined to be the long-run average intake of the nutrient, and is unobservable. Nusser *et al* (1992) have presented a procedure for estimating the usual intake distribution from data obtained on more than one day for at least some individuals in the sample. It is assumed that the observations are independent. The procedure consists of transforming observed intake data into normal random variables, estimating the parameters of the usual intake distribution from a measurement error model, and obtaining the estimated usual intake distribution in the original scale by applying standard results on the distribution of transformed random variables. An and Carriquiry (1991) extended those results to the case of nonindependent days of intake data.

The objective of this work was to obtain estimates for the standard errors of the estimated quantiles of the estimated distributions of usual nutrient intake. The estimation of the standard errors is different because quantiles are estimated from transformed data and these estimated quantiles depend on estimated variance components.

Data from the 1985 Continuing Survey of Food Intakes by Individuals (CSFII) has been used to estimate quantiles of the distribution of usual intakes for six selected dietary components. The 1985 CSFII data were collected by a contractor for the Human Nutrition Information Service (HNIS) of the USDA from April 1985 through March 1986. The 1985 CSFII data consists of twenty-four-hour nutrient intake recall data collected by interview for the first day, and recall data for subsequent days (3 more days) collected by telephone whenever possible. Data were collected from 19 to 50 year-old women and from their children aged 1 to 5 years. Intakes were obtained at no less than 2 months intervals over the one-year period (USDA, 1987). A subset of the CSFII four-day data set, consisting of 25 to 50 year-old women who were not pregnant or lactating and who were main meal planners or preparers, was used in the analysis. These individuals were selected to provide a relatively homogeneous group for analysis and to match the Recommended Dietary Allowances classification (National Research Council, 1989). The estimated quantiles of the usual intake distributions and their standard errors based on the methodology described in this paper are presented.

2 ESTIMATION OF VARIANCE COMPONENTS

Let Y_{ij} represent the observed intake of a nutrient for the i-th individual on the j-th day ($i = 1, ...m$; $j =$

*This research was partly supported by Research Support Agreement 58-3198-9-032 with the Human Nutrition Information Service, U.S. Department of Agriculture.

Table 1 **Analysis of Variance for Random Components Model**

Source	df	MS	EMS
Between Individuals	d_B	MSB	$\sigma_u^2 + k_0\sigma_x^2$
Within Individuals	d_W	MSW	σ_u^2

$k_0 = (m-1)^{-1}(n - n^{-1}\sum_{i=1}^m k_i^2)$
$n = \sum_{i=1}^m k_i$.
$d_B = m - 1$
$MSB = d_B^{-1}\sum_{i=1}^m k_i(\overline{X}_{i\cdot} - \overline{X}_{\cdot\cdot})^2$
$d_W = \sum_{i=1}^m (k_i - 1)$
$MSW = d_W^{-1}\sum_{i=1}^m \sum_{j=1}^{k_i}(X_{ij} - \overline{X}_{i\cdot})^2$

$1, ..., k_i$). Let $g(\cdot)$ denote the semiparametric transformation presented by Nusser *et al* (1992), and thus $g(Y_{ij}) = X_{ij}$, where the X_{ij} are normally distributed random variables. Furthermore, assume that

$$X_{ij} = x_i + u_{ij}, \qquad i = 1, ..., m;\; j = 1, ..., k_i \quad (1)$$

where $x_i \sim \text{NI}(\mu, \sigma_x^2)$, $u_{ij} \sim \text{NI}(0, \sigma_u^2)$, and u_{ij} is independent of x_k for all i, j, k. In model (1), X_{ij} represents the transformed observed usual intake of the i-th individual on the j-th interview day for a nutrient, x_i represents the true usual intake in normal scale of the i-th individual, and u_{ij} represents the measurement error associated with the i-th individual on the j-th interview day . Unknown variance components σ_x^2 and σ_u^2 can be estimated from the appropriate mean squares in the analysis of variance (ANOVA) table presented in Table 1. (Henderson, 1953).

Estimators associated with the table are:

$$\begin{aligned}
\hat{\mu} &= \overline{X}_{\cdot\cdot} = \frac{1}{m}\sum_{i=1}^m \overline{X}_{i\cdot} = \frac{1}{m}\sum_{i=1}^m\left[\frac{1}{k_i}\sum_{j=1}^{k_i} X_{ij}\right], \\
\widehat{\sigma_u^2} &= MSW = \frac{1}{d_W}\sum_{i=1}^m\sum_{j=1}^{k_i}(X_{ij} - \overline{X}_{i\cdot})^2, \\
\widehat{\sigma_x^2} &= \frac{1}{k_0}(MSB - MSW) \\
&= \frac{1}{k_0 d_B}\sum_{i=1}^m k_i(\overline{X}_{i\cdot} - \overline{X}_{\cdot\cdot})^2 \\
&\quad - \frac{1}{k_0 d_W}\sum_{i=1}^m\sum_{j=1}^{k_i}(X_{ij} - \overline{X}_{i\cdot})^2
\end{aligned} \quad (2)$$

where k_0 is the coefficient for σ_x^2 in the between-individuals expected mean squares (Snedecor and Cochran, 1989, pp245), and d_W and d_B are the within- and between-individuals degrees of freedom respectively. For the i-th individual, define

$$\begin{aligned}
Z_{i1} &= \overline{X}_{i\cdot} = \frac{1}{k_i}\sum_{j=1}^{k_i} X_{ij}, \\
Z_{i2} &= \frac{m}{m-1}\cdot\frac{k_i}{k_0}(\overline{X}_{i\cdot} - \overline{X}_{\cdot\cdot})^2, \\
Z_{i3} &= \frac{m}{m-1}\cdot\frac{1}{(k_i - 1)k_0}\sum_{j=1}^{k_i}(X_{ij} - \overline{X}_{i\cdot})^2,
\end{aligned} \quad (3)$$

thus $\underline{Z}_i = (Z_{i1}, Z_{i2}, Z_{i3})'$ $(i = 1, ..., m)$ are a sample from a distribution with the covariance matrix V_{ZZ}. Therefore, estimators in (2) can be considered as estimators using the sample $\{\underline{Z}_i,\ i = 1, ..., m\}$. Furthermore, if $k_i \equiv k_0$ for all i, then

$$\begin{aligned}
\hat{\mu} &= \overline{Z}_{\cdot 1}, \\
\widehat{\sigma_u^2} &= k_0\overline{Z}_{\cdot 3}, \\
\widehat{\sigma_x^2} &= \overline{Z}_{\cdot 2} - \overline{Z}_{\cdot 3},
\end{aligned} \quad (4)$$

where $\overline{Z}_{\cdot j} = \sum_{i=1}^m Z_{ij}/m$, $(j = 1, 2, 3)$.

The covariance matrix V_{ZZ} is estimated by ordinary sampling formulas using PC CARP (PC CARP, 1989). Let the estimated covariance matrix be

$$\widehat{V}_{ZZ} = \begin{pmatrix} v_{11} & v_{12} & v_{13} \\ v_{21} & v_{22} & v_{23} \\ v_{31} & v_{32} & v_{33} \end{pmatrix} \quad (5)$$

Then the estimators of the variances for $\hat{\mu}$ and $\hat{\sigma}_x^2$ are

$$\begin{aligned}
\widehat{V}(\hat{\mu}) &= v_{11}, \\
\widehat{V}(\hat{\sigma}_x^2) &= v_{22} + v_{33} - 2v_{23}.
\end{aligned} \quad (6)$$

Furthermore, the Taylor expansion of

$$\hat{\sigma}_x = \sqrt{\widehat{\sigma_x^2}} \quad (7)$$

around σ_x^2 gives

$$\hat{\sigma}_x \cong \sigma_x + \frac{1}{\sqrt{4\sigma_x^2}}\left(\hat{\sigma}_x^2 - \sigma_x^2\right). \quad (8)$$

Therefore, an estimator of the variance for $\hat{\sigma}_x$ can be obtained from (4) and (6) as

$$\widehat{V}(\widehat{\sigma_x}) = \frac{1}{4\hat{\sigma}_x^2}\widehat{V}(\hat{\sigma}_x^2). \quad (9)$$

Estimators (4) and (6) will be used in the estimated standard errors.

3 ESTIMATION OF QUANTILES

Suppose $x \sim N(\mu, \sigma_x^2)$. Given $p \in (0,1)$, let Q_p^x denote the p-percentile of x, i.e. Q_p^x is the quantile such that

$$F(Q_p^x) = \Pr(x \leq Q_p^x) = \Phi\left(\frac{Q_p^x - \mu}{\sigma_x}\right) = p.$$

where Φ is the cumulative distribution function of a N(0,1) random variable. Then an estimator of the quantile Q_p^x using (2) and (7) is

$$\widehat{Q_p^x} = \widehat{\sigma}_x \Phi^{-1}(p) + \widehat{\mu}. \quad (10)$$

In order to obtain the variance of $\widehat{Q_p^x}$, using (8), we write $\widehat{Q_p^x}$ as

$$\begin{aligned} \widehat{Q_p^x} &= \Phi^{-1}(p)\left[\sigma_x - \frac{1}{2\sigma_x}(\sigma_x^2 - \widehat{\sigma}_x^2)\right] + \widehat{\mu} \\ &= \frac{\Phi^{-1}(p)\sigma_x}{2} + \overline{Z}_{\cdot 1} \\ &\quad + \frac{\Phi^{-1}(p)}{2\sigma_x}\left(\overline{Z}_{\cdot 2} - \overline{Z}_{\cdot 3}\right) \quad (11) \\ &= \left(1, \frac{\Phi^{-1}(p)}{2\sigma_x}, -\frac{\Phi^{-1}(p)}{2\sigma_x}\right)\begin{pmatrix} \overline{Z}_{\cdot 1} \\ \overline{Z}_{\cdot 2} \\ \overline{Z}_{\cdot 3} \end{pmatrix} \\ &\quad + \frac{\Phi^{-1}(p)\sigma_x}{2}, \end{aligned}$$

therefore

$$V\left(\widehat{Q_p^x}\right) \doteq \underline{b}' V_{ZZ}\, \underline{b}, \quad (12)$$

where $\underline{b}' = \left(1, \Phi^{-1}(p)/(2\sigma_x), -\Phi^{-1}(p)/(2\sigma_x)\right)$. An estimator for $V(\widehat{Q_p^x})$,

$$\widehat{V}\left(\widehat{Q_p^x}\right) = \widehat{\underline{b}}' \widehat{V}_{ZZ} \widehat{\underline{b}} \quad (13)$$

where $\widehat{\underline{b}}' = \left(1, \Phi^{-1}(p)/(2\widehat{\sigma}_x), -\Phi^{-1}(p)/(2\widehat{\sigma}_x)\right)$ and $\widehat{V}_{ZZ}$ is defined in (5).

Let the random variable W be a fixed known function of $x \sim N(\mu, \sigma_x^2)$,

$$W = h(x) \quad (14)$$

where h has continuous first and second derivatives. Let Q_p^W denote the p-percentile of W. Then $Q_p^W = h(Q_p^x)$, and an estimator for Q_p^W is given by

$$\widehat{Q_p^W} = h\left(\widehat{Q_p^x}\right). \quad (15)$$

Table 2 **Basic Sample Features of the 1985 CSFII data**

Nutrient	Units	mean	S.E.	Skewness coefficient
Calcium	mg	624	303	1.19
Energy	kcal	1,672	552	0.65
Iron	mg	11.1	4.1	1.25
Protein	g	65.8	21.8	0.78
Vitamin A	μg RE	839	971	6.79
Vitamin C	mg	79	52	1.24

For the 1985 CSFII data, usual intake quantiles are estimated by (15). In our analysis we treat $h(\cdot)$ as a fixed known function. To derive an expression for the variance of $\widehat{Q_p^W}$, expand $\widehat{Q_p^W}$ in a Taylor series around Q_p^x to obtain

$$\widehat{Q_p^W} \cong h\left(Q_p^x\right) + h'\left(Q_p^x\right)\left[\widehat{Q_p^x} - Q_p^x\right], \quad (16)$$

where h' is the derivative of h, and estimate $V(\widehat{Q_p^W})$ as

$$\widehat{V}\left(\widehat{Q_p^W}\right) = \left[h'\left(Q_p^x\right)\right]^2 \left[\widehat{V}\left(\widehat{Q_p^x}\right)\right], \quad (17)$$

where $\widehat{V}\left(\widehat{Q_p^x}\right)$ is as in (13).

4 NUMERICAL RESULTS

We applied our results to the 1985 CSFII data to estimate quantiles of the usual intakes for six selected nutrients: calcium, energy, iron, protein, vitamin A and vitamin C. Basic features of the distributions of the six dietary components in the 1985 CSFII data were obtained by calculating the mean, standard error, and skewness coefficient. These estimates are presented in Table 2 (Carriquiry *et al*, 1992). The estimated skewness coefficients indicate that the distributions of usual intake are skewed to the right for all of the nutrients considered. The within- and between-individual variances of the transformed observed intakes are given in Table 3 for each of six nutrients. Using the methods presented in the previous section, the estimated quantiles and the standard errors in the original scale are shown in Table 4 for each of six nutrients.

5 REFERENCES

[1] An, A. B. and Carriquiry, A. L. (1991). Estimation of the correlation among days for transformed nutrient intake.

[2] Anderson, (1958). An Introduction to Multivariate Statistical Analysis. Wiley, New York.

Table 3 **Estimated Within- and Between-individual Variances of Transformed Observed Usual Intakes for Each of Six Nutrients**

Nutrient	Within-individual variance $\hat{\sigma}_u^2$	Between-individual variance $\hat{\sigma}_x^2$
Calcium	0.634	0.366
Energy	0.626	0.375
Iron	0.685	0.315
Protein	0.726	0.274
Vitamin A	0.745	0.255
Vitamin C	0.682	0.318

[3] Carriquiry, A. L. , Helen, J. H., Dodd, K. W., Nusser, S. M., Guenther, P.M. and Fuller, W.A. (1992). Estimating usual nutrient intake distributions. Journal Paper No. J-14654 of the Iowa Agriculture and Home Economics Experiment Station, Ames, IA.

[4] Fuller, W.A. (1987). Measurement Error Models. Wiley, New York, NY.

[5] Nation Research Council (1989). Recommended Dietary Allowances, 10th ed. National Academy Press, Washington, DC.

[6] Nusser, S. M., Carriquiry, A. L. and Fuller, W. A. (1992). A semiparametric approach to estimating usual intake distributions. CARD working paper 92-WP99, Iowa State University, Ames, IA..

[7] PC CARP (1989). Statistical Laboratory, Iowa State University. Ames, IA.

[8] Snedecor, G.W. and Cochran, W. G. (1989). Statistical Methods, 8th ed. Iowa State University Press, Ames, IA.

[9] U.S. Department of Agriculture, Human Nutrition Information Service, Nutrition Monitoring Division. Women 19-50 years and their children, 1-5 years, 4 days. National Food Consumption Survey Continuing Survey of Food Intakes by Individuals. NFCS, CSFII: Report No. 85-4, 1987.

Table 4 **Estimated Usual Intake Quantiles and Their Standard Errors for Each of Six Nutrients**

Percent	Calcium	Energy	Iron
0.010	215.46	790.45	5.28
	(12.35)	(32.52)	(0.26)
0.025	258.14	904.79	6.02
	(12.36)	(31.68)	(0.25)
0.050	299.34	1009.14	6.69
	(12.28)	(30.83)	(0.23)
0.100	352.38	1136.35	7.50
	(12.14)	(29.82)	(0.22)
0.250	455.49	1365.33	8.96
	(12.34)	(28.50)	(0.20)
0.500	593.70	1643.77	10.78
	(14.67)	(28.69)	(0.21)
0.750	759.91	1947.39	12.91
	(20.82)	(31.96)	(0.26)
0.900	936.14	2242.18	15.24
	(29.95)	(38.06)	(0.36)
0.950	1054.73	2428.34	16.91
	(37.04)	(43.07)	(0.46)
0.975	1166.04	2595.73	18.56
	(44.18)	(48.16)	(0.58)
0.990	1305.81	2797.26	20.76
	(53.67)	(54.86)	(0.75)

Percent	Protein	Vitamin A	Vitamin C
0.010	34.25	226.91	18.31
	(1.49)	(16.60)	(1.39)
0.025	38.53	273.26	23.35
	(1.40)	(17.14)	(1.48)
0.050	42.36	319.00	28.46
	(1.32)	(17.69)	(1.55)
0.100	46.92	380.00	35.37
	(1.24)	(18.51)	(1.64)
0.250	54.97	508.30	49.90
	(1.15)	(21.07)	(1.89)
0.500	64.66	709.26	71.66
	(1.19)	(28.34)	(2.54)
0.750	75.39	1007.21	100.28
	(1.45)	(46.50)	(3.81)
0.900	86.23	1408.19	131.93
	(1.91)	(82.21)	(5.50)
0.950	93.43	1739.99	153.33
	(2.29)	(118.97)	(6.71)
0.975	100.16	2105.96	173.23
	(2.69)	(164.89)	(7.85)
0.990	108.59	2651.83	197.87
	(3.23)	(241.22)	(9.36)

Numbers in brackets are estimated standard errors.