Estimated Correlations Among Days for the Combined 1989-91 CSFII

Dietary Assessment Research Series Report 4

A.L. Carriquiry, W.A. Fuller, J.J. Goyeneche, and H.H. Jensen

Staff Report 95-SR 77
September 1995

Estimated Correlations Among Days for the Combined 1989-91 CSFII

Dietary Assessment Research Series Report 4

A.L. Carriquiry, W.A. Fuller, J.J. Goyeneche, and H.H. Jensen

Staff Report 95-SR 77
September 1995

Center for Agricultural and Rural Development
Iowa State University
Ames, IA 50011

A.L. Carriquiry is associate professor, W.A. Fuller is distinguished professor, J.J. Goyeneche is a graduate research assistant, in the statistics department, Iowa State University. H.H. Jensen is associate professor of economics, Iowa State University, and head of the Food and Nutrition Policy Division, Center for Agricultural and Rural Development.

Partial funding support was provided under Cooperative Research Agreement #58-3198-2-006 between the Agricultural Research Service, USDA, and Iowa State University.

CONTENTS

TABLES

ABSTRACT

Data obtained from dietary intake surveys are often used to estimate the proportion of the population with insufficient (or excessive) intake of certain dietary components. It is generally agreed that the usual or long-run average intake of a nutrient is the appropriate measure of an individual's intake. In this light, assessments of the dietary status of the population should be based on the distribution of usual intakes for each dietary component.

Several methods have been proposed for estimating the distributions of usual intakes from dietary data. The methodology presented by Nusser et al. (1995) is very appealing, since it accounts for all of the attributes of dietary intake data, and explicitly recognizes that when data are collected on consecutive days, it is necessary to incorporate the correlations among intake days into the estimation procedure.

In this paper we present a method for obtaining smooth estimates of day-to-day correlations for nutrient intake, and their standard errors, when dietary data are collected on consecutive days. The method consists of two steps. In the first step, direct correlation estimates are obtained by fitting a measurement error model with a second-order autoregressive variance structure. Smoothed correlation estimates are then obtained from the direct estimates by means of a two-factor model.

Data from the combined 1989-CSFII are used to obtain correlation estimates (and their standard errors) for 26 nutrients, water, and six ratios of dietary components. Estimates were obtained for each of five age-sex groups.

ESTIMATED CORRELATIONS AMONG DAYS FOR THE COMBINED 1989-91 CSFII

Data obtained from dietary intake surveys are often used to estimate the proportion of the population with insufficient (or excessive) intake of certain dietary components. It is generally recognized that an individual who has a low intake of a dietary component on one day is not necessarily deficient for that dietary component. Rather, it is low intake over an extended period that causes a dietary deficiency. It is therefore the long-run average intake–or usual intake–of dietary components that is the appropriate measure of an individual's intake. In this light, assessments of the dietary status of the population must be based on the distribution of usual intakes for each dietary component.

Several approaches have been proposed for estimating the distribution of usual intakes from dietary data (e.g., National Academy of Sciences, 1986; Nusser et al., 1995). An implicit assumption in the methodology proposed in the NAS report is that intakes observed for each individual in the sample are independent, both across individuals and across days within an individual. The approach proposed by Nusser et al., that is referred to as the Iowa State University (ISU) method, explicitly recognizes that intake data that are obtained on consecutive days may violate the independence assumption required by the procedure outlined in the NAS report. Given that dietary data are often obtained from a sample of individuals on consecutive days, the ISU method allows for the incorporation of the correlation among intake days when estimating usual intake distributions. It is thus necessary to estimate the day-to-day correlation for each dietary component under study, and for appropriate age-sex groups.

Dietary intake data from the combined 1989, 1990, and 1991 Continuing Survey of Food Intake by Individuals (CSFII) were used to estimate the correlation among days for 27 dietary components, water, and six ratios of dietary components (e.g., percent, calories from total fat). Data were collected on three consecutive days for each individual in the sample, during 1989, 1990 and 1991. Data used to estimate the correlations were obtained from 11,912 individuals: 2,494 men aged 20-59, 3,168 women aged 20-59, 887 males over 59 years, 1,451 females over 59 years, and 3,710 individuals under 20 years of age. The surveys were conducted by the Agricultural Research Service (ARS) of the U.S. Department of Agriculture.

In this report, we describe a two-step procedure to estimate the day-to-day correlation coefficients, and present the results that were obtained. The first step in this approach consists in obtaining direct estimates of the correlation coefficients by following the procedure outlined in An and Carriquiry (1991). A description of the procedure is presented here. In the second step, smoothed correlation estimates are obtained by fitting a measurement error model to the direct estimates of the correlation coefficients. Results obtained from applying the method to the combined 1989-91 CSFII intake data are presented. A method for obtaining nearly unbiased estimates of the ratio of two dietary components for an individual on any day is proposed in Carriquiry et al. (1995).

Methods

Model

Let Y_{ij} represent the observed intake of a dietary component (or a ratio) for individual i on day $j\,(j = 1,\,2,\,3)$ in a given age-sex group. Here, Y_{ij} represents an observed intake that has been ratio adjusted by survey related effects (Nusser et al. 1994). Let ρ represent the correlation among intake days for a dietary component. To obtain a direct estimate of ρ , as will be seen later, individual effects on intakes are removed by fitting a linear model to observed intakes. In this linear model formulation, it is assumed that observed intakes, as well as measurement errors, are normally distributed random variables. As has been noted earlier, however, (NAS 1986; Battese et al. 1988; Nusser et al. 1995), observed dietary intakes are not normally distributed, suggesting that (unobservable) usual intakes are not normally distributed either. Nusser et al. (1995), extending the methodology proposed in the NAS report (1986), suggested that dietary intakes be transformed into normal random variables prior to analysis. They proposed a two-step transformation method that consists of the following. First, using a procedure such as the one suggested by Lin and Vonesh (1989) obtain the "best" power transformation for the dietary intake data. Here, "best" refers to that power α such that the distribution of the Y_{ij}^{α} is as close to normal as possible. Second, apply a nonparametric method (cubic splines) to the power transformed observations to obtain a set of transformed observations X_{ij} that are approximately normally distributed. In this work, we have applied only the power transformation to the observed intakes. That is, correlation coefficients were estimated from power transformed intakes whose distribution approached, but was not necessarily, normal. The rationale was that the correlation estimates were based on a linear model and functions of suitably chosen linear contrasts; it is known that such approaches are robust to slight departures from the normality assumption (e.g., Neter et al. 1990).

Dietary data were collected from a multi-stage stratified area probability sample from the conterminous 48 states. While the sample was designed to be self-weighting for several important characteristics, not all individuals in the original sample provided a response for all three survey days. Weights were therefore constructed to correct for nonresponse in the three-day sample. For a description of the weighting procedure and a list of the weights computed for the 1989-91 CSFII, refer to An et al. (1994). We examined the effect of incorporating the sampling weights on the estimated correlation coefficients and their standard errors, and found them to be negligible. In this work, therefore, estimates of the correlation coefficients were computed on the unweighted intake data.

In what follows, X_{ij} denotes the transformed intake for the i-th individual on the j-th day, $i = 1, 2, ..., n$, $j = 1, 2, 3$. Consider the linear model

$$\begin{aligned} X_{ij} &= \mu + \gamma_i + \epsilon_{ij}, \\ E(\gamma_i) &= 0, \, \epsilon_{ij} \sim N\left(0, \sigma_\epsilon^2\right), \end{aligned} \tag{1}$$

where γ_i represents the individual effect on intake. The linear model above states that the dietary intake observed for an individual on any given day is a linear combination of a general effect μ, an individual effect γ_i, and a measurement error ϵ_{ij}. Furthermore, assume

$$\begin{aligned} E\left(\epsilon_{ij},\, \epsilon_{i',j+k}\right) &= 0 && \text{if } i \neq i' \\ &= \sigma_\epsilon^2 && \text{if } i = i' \text{ and } k = 0 \\ &= \rho\sigma_\epsilon^2 && \text{if } i = i' \text{ and } k = 1 \\ &= \rho^2\sigma_\epsilon^2 && \text{if } i = i' \text{ and } k = 2. \end{aligned}$$

Direct Estimation of ρ

Our objective is to estimate the parameter ρ. Consider the model described in the preceding section and the linear contrasts

$$\begin{aligned} L_{1i} &= X_{i1} - X_{i3} = \mu - \mu + \gamma_i - \gamma_i + \epsilon_{i1} - \epsilon_{i3} = \epsilon_{i1} - \epsilon_{i3}, \\ L_{2i} &= X_{i1} - 2X_{i2} + X_{i3} = \epsilon_{i1} - 2\epsilon_{i2} + \epsilon_{i3}, \end{aligned}$$

among the day effects. The contrast L_1 reflects the linear day effect on intake for an individual, while L_2 represents second order day effects. Under model (1), the variance-covariance matrix of the joint distribution of L_1 and L_2 is

$$\begin{pmatrix} 2\left(1-\rho^2\right) & 0 \\ 0 & 6 - 8\rho + 2\rho^2 \end{pmatrix} \sigma_\epsilon^2,$$

which depends on the unknown correlation coefficient ρ as well as on σ_ϵ^2. Therefore, the ratio

$$\frac{\mathrm{Var}\,(L_1)}{\mathrm{Var}\,(L_2)} = \frac{\left(1-\rho^2\right)}{(1-\rho)^2 + 2\,(1-\rho)} = \frac{1+\rho}{3-\rho}$$

depends only on the parameter ρ. Thus, an expression for ρ can be obtained from the ratio of the following function of contrast variances:

$$\rho = \frac{3\,\mathrm{Var}\,(L_1) - \mathrm{Var}\,(L_2)}{\mathrm{Var}\,(L_1) + \mathrm{Var}\,(L_2)}.$$

An estimate for ρ can be obtained by replacing the unknown contrast variances with their estimators

$$\hat{\rho} = \frac{3S_1^2(L_1) - S_2^2(L_2)}{S_1^2(L_1) + S_2^2(L_2)}, \tag{2}$$

where $S_1^2(L_1)$ and $S_2^2(L_2)$ can be computed from the transformed sample values X_{ij} as

$$S_k^2(L_k) = \frac{1}{n-1}\sum_{i=1}^{n}(L_{ki} - \bar{L}_{k.})^2, \quad k = 1, 2.$$

Here, $\bar{L}_{k.} = n^{-1}\sum_{i=1}^{n} L_{ki}$ is the mean of the k-th contrast computed over the n individuals in the sample.

We now derive an expression for the standard deviation of $\hat{\rho}$. Under model (1),

$$\begin{pmatrix} L_1 \\ L_2 \end{pmatrix} \sim N\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, 2(1-\rho)\sigma_\epsilon^2 \begin{pmatrix} 1+\rho & 0 \\ 0 & 3-\rho \end{pmatrix}\right).$$

The variance estimators S_1^2, S_2^2 are independently distributed. Then, for $k = 1, 2$,

$$E\left(S_k^2\right) = \sigma_k^2,$$

and

$$\text{Var}\left(S_k^2\right) = \frac{2\left(\sigma_k^2\right)^2}{n-1},$$

since under the assumptions of model (1), $\frac{n-1}{\sigma_k^2}S_k^2$ is distributed as a χ^2 random variable with $n-1$ degrees of freedom. To obtain an expression for the standard error of $\hat{\rho}$, we expand $\hat{\rho}$ around (σ_1^2, σ_2^2), the true contrast variances, using Taylor's Theorem. Recall from (2) that

$$\hat{\rho} = f\left(S_1^2, S_2^2\right) = \frac{3S_1^2 - S_2^2}{S_1^2 + S_2^2}. \tag{3}$$

Then, the Taylor Series expansion of (1) around (σ_1^2, σ_2^2) is

$$\hat{\rho} - \rho \doteq f\left(S_1^2, S_2^2\right)|_{(\sigma_1^2, \sigma_2^2)} - f\left(\sigma_1^2, \sigma_2^2\right) + [\lambda_1\ \lambda_2]\left(S_1^2 - \sigma_1^2, S_2^2 - \sigma_2^2\right)', \tag{4}$$

where $\boldsymbol{\lambda}' = [\lambda_1\ \lambda_2]$ is the gradient vector of $\hat{\rho} = f\left(S_1^2\ S_2^2\right)$. Thus, $\boldsymbol{\lambda}$ is the vector of first order partial derivatives of $\hat{\rho}$ with respect to S_1^2 and S_2^2, evaluated at the point $\left(\sigma_1^2, \sigma_2^2\right)$. The first order partial derivatives of $\hat{\rho} = f\left(S_1^2,\, S_2^2\right)$ with respect to S_1^2 and S_2^2 are given by:

$$\begin{aligned} \frac{\partial}{\partial S_1^2}\hat{\rho} &= \frac{4S_2^2}{\left(S_1^2+S_2^2\right)^2}, \\ \frac{\partial}{\partial S_2^2}\hat{\rho} &= -\frac{4S_1^2}{\left(S_1^2+S_2^2\right)^2}. \end{aligned}$$

Therefore, the gradient vector $\boldsymbol{\lambda}$, with derivatives evaluated at $\left(\sigma_1^2,\, \sigma_2^2\right)$ is given by

$$\boldsymbol{\lambda}' = \left(\frac{4\sigma_2^2}{\left(\sigma_1^2+\sigma_2^2\right)^2},\, -\frac{4\sigma_1^2}{\left(\sigma_1^2+\sigma_2^2\right)^2}\right).$$

Expression (4) can therefore be written as

$$\hat{\rho} - \rho = \boldsymbol{\lambda}'\left[S_1^2 - \sigma_1^2,\; S_2^2 - \sigma_2^2\right]',$$

and the variance of $\hat{\rho}$ can be approximated by

$$\begin{aligned} \hat{\sigma}^2\left(\hat{\rho}\right) &= \hat{E}\left(\hat{\rho}-\rho\right)^2 = \boldsymbol{\lambda}'\text{var}\left\{\left(S_1^2-\sigma_1^2,\; S_2^2-\sigma_2^2\right)'\right\}\boldsymbol{\lambda} \\ &= \boldsymbol{\lambda}'\begin{pmatrix} \frac{2\left(\sigma_1^2\right)^2}{n-1} & 0 \\ 0 & \frac{2\left(\sigma_2^2\right)^2}{n-1} \end{pmatrix}\boldsymbol{\lambda} \qquad (5) \\ &= \frac{\left(1+\rho\right)^2\left(3-\rho\right)^2}{4\left(n-1\right)}, \end{aligned}$$

where $\sigma_1^2 = 2\left(1-\rho\right)\left(1+\rho\right)\sigma_\epsilon^2$, and $\sigma_2^2 = 2\left(1-\rho\right)\left(3-\rho\right)\sigma_\epsilon^2$. An estimator for the standard deviation of $\hat{\rho}$ can then be obtained by evaluating (5) at $\rho = \hat{\rho}$,

$$\hat{\sigma}\left(\hat{\rho}\right) = \sqrt{\frac{\left(1+\hat{\rho}\right)^2\left(3-\hat{\rho}\right)^2}{4\left(n-1\right)}} \qquad (6)$$

It can be shown that this estimator is a consistent estimator of $\sigma\left(\hat{\rho}\right)$.

Model Estimation for Correlations

Let $\hat{\rho}_{k\ell}$ denote the estimated day-to-day correlation for the k-th dietary component (or ratio of dietary components) in the ℓ-th age-sex group, obtained as described in the preceding section.

Theoretical considerations as well as empirical evidence suggest that the correlation estimates for a particular dietary component are similar for the different age-sex groups. Furthermore, the variation among group mean correlations is small relative to the size of the estimation error. This suggests that the variation in correlations can be explained by a model that contains parameters for the age-sex group and parameters for the dietary components. We use a measurement error model where the vector of five estimated correlations for a dietary component is the vector of observations.

The vector of unknown true correlations is assumed to depend on a vector of factors of dimensions smaller than five that represent the age-sex effects. The analysis of the Appendix indicates that two factors are sufficient to explain the variability among the five age-sex groups. We write the model as

$$\begin{aligned} \rho_{k\ell} &= \mu_\ell + \beta_{1\ell}(\rho_{k4} - \mu_4) + \beta_{2\ell}(\rho_{k5} - \mu_5), \quad \ell = 1, \ldots, 5, \\ \hat{\rho}_{k\ell} &= \rho_{k\ell} + e_{k\ell}, \end{aligned} \tag{7}$$

where $\hat{\rho}_{k\ell}$ is the estimated day-to-day correlation for the k-th component for the ℓ-th group, μ_ℓ is the population mean over components for the ℓ-th group, $\rho_{k\ell}$ is the unobservable true day-to-day correlation for the k-th component and ℓ-th group, $e_{k\ell}$ is the estimation error associated with $\hat{\rho}_{k\ell}$ and $\beta_{i\ell}$, $i = 1,\ 2,\ \ell = 1, 2, 3$ are parameters to be estimated. In the suggested parameterization, $\beta_{14} = \beta_{25} = 1$, $\beta_{24} = \beta_{15} = 0$. That is, the true values for the first three age-sex groups are expressed as a linear function of the true values of the last two groups. This is an arbitrary specification and other parameterizations produce the same final estimates of the true day-to-day correlations. In the estimation, our working assumption is

$$(e_{k1},\, e_{k2},\, e_{k3},\, e_{k4},\, e_{k5})' = NI\left(0, \Sigma_{ee}\right),$$

where Σ_{ee} is a diagonal matrix. The elements of Σ_{ee} were estimated by the average estimated variance (over all dietary components) $C^{-1} \sum_{j=1}^{C} \hat{\sigma}^2\left(\hat{\rho}_{j\ell}\right)$, where $C = 34$ is the total number of components. The error distribution working assumption is an approximation because it is reasonable to believe that the errors in component estimates within a group are correlated. Also, the assumption of normality is not required and the evidence suggests that direct estimates of day-to-day correlations are not normally distributed. Estimation for this model is discussed, for example, by Fuller (1987, Section 4.1).

Results

The procedure described above was used to compute correlations among days for each of 27 dietary components, water, and six ratios in each of the five sex-age groups, Dietary data were obtained from the 1989, 1990, and 1991 CSFII, and the analysis was performed on observations from all three years combined. In the case of ratios, "observations" were computed as proposed in Carriquiry et al. (1995).

Observed intakes were transformed within each age-sex group prior to analysis. The transformation consisted in a nonlinear (power) transformation to correct for nonnormality of the observations. Best estimated powers for each dietary component and each sex-age group are presented in Table 1. Direct estimates of the correlation coefficients are shown in Table 2 for all dietary components and age-sex groups. We have adopted the following notation for the six ratios of dietary components. Percent fat denotes the percent of total calories consumed from total fat. Percent mfat, percent sfat, and percent pfat represent calories from monounsaturated, saturated, and polyunsaturated fat, respectively. Finally, calories from carbohydrates and protein are denoted percent carb and percent prot, respectively. Expression (6) was used to estimate the standard errors of the correlation coefficients. Average values (over all dietary components) of $\hat{\sigma}\ (\hat{\rho})$ were 0.0319, 0.0547, 0.0283, 0.0424, and 0.0257 for males 20-59, males over 59, females 20-59, females over 59, and individuals 19 or younger, respectively. The majority of the estimated correlation coefficients are significantly different from zero at the $\alpha = 0.05$ level.

The estimates for the five groups are positively correlated. Correlation estimates among the day-to-day correlations are given in Table 3. A procedure for estimating the correlations in Table 3 is presented in Appendix A. Details can be found in Fuller (1987, Section 4.1).

The estimates of the day-to-day correlations constructed under the model are given in Table 4. The estimates of the parameters of model (7) are given in the appendix. Using estimates of the model parameters, it is possible to construct improved estimates of the correlation for each cell of the table. The estimates in Table 4 are somewhat analogous to the predicted values one might obtain from a regression model.

Table 1. Inverse of power used for transformation

Component	Males 20-59	Males ≥ 60	Females 20-59	Females ≥ 60	Persons 0-19
Calcium	5.5	2.5	3.5	4.0	2.0
Carbohydrates	3.0	1.5	2.5	2.5	2.0
Carotene	0.0	0.0	0.0	0.0	0.0
Cholesterol	4.0	3.0	4.0	4.5	3.0
Copper	10.0	8.0	4.0	6.5	2.5
Fiber	3.5	2.5	2.5	3.0	2.5
Folate	8.5	5.0	4.5	4.5	3.0
Iron	10.0	4.5	6.5	10.0	3.0
Energy	3.5	1.5	2.5	2.5	2.0
Magnesium	3.5	2.0	2.5	2.5	2.0
Monounsaturated fat	3.0	2.5	2.5	3.5	2.0
Niacin	3.5	2.5	2.5	3.0	2.0
Polyunsaturated fat	4.0	3.0	3.5	4.0	3.5
Phosphorus	3.5	2.0	2.5	3.0	2.0
Potassium	3.0	1.5	2.0	2.0	2.0
Protein	3.0	2.0	2.0	2.0	2.0
Riboflavin	6.0	3.0	3.0	4.0	2.0
Saturated fat	3.0	2.5	2.5	3.5	2.0
Sodium	3.0	2.5	3.0	3.5	2.0
Total fat	3.0	2.0	2.5	3.0	2.5
Thiamin	4.5	3.0	3.0	3.5	2.0
Vitamin B6	4.5	2.5	3.0	3.0	2.5
Vitamin B12	10.0	10.0	7.0	10.0	4.0
Vitamin A	9.5	7.0	7.5	10.0	3.5
Vitamin C	4.5	3.5	3.5	3.0	4.5
Vitamin E	10.0	10.0	6.5	10.0	4.5
Water	4.0	1.5	2.5	2.0	2.0
Zinc	0.0	0.0	4.0	7.0	2.5

Table 2. Direct estimates of day-to-day correlations

Component	Males 20-59	Males ≥ 60	Females 20-59	Females ≥ 60	Persons 0-19
Calcium	0.100	0.084	0.135	0.154	0.116
Carbohydrates	0.092	0.180	0.133	0.215	0.187
Carotene	0.077	0.072	0.030	0.115	0.068
Cholesterol	0.059	0.006	0.036	-0.054	0.030
Copper	0.118	0.136	0.125	0.203	0.106
Fiber	0.073	0.150	0.098	0.277	0.113
Folate	0.144	0.151	0.131	0.088	0.085
Iron	0.055	0.226	0.081	0.152	0.026
Energy	0.133	0.247	0.083	0.172	0.146
Magnesium	0.132	0.239	0.181	0.269	0.117
Monounsaturated fat	0.126	0.199	0.048	0.007	0.079
Niacin	0.104	0.147	0.062	0.194	-0.018
Polyunsaturated fat	0.066	0.141	0.067	0.045	0.052
Phosphorus	0.110	0.201	0.107	0.236	0.071
Potassium	0.098	0.217	0.175	0.227	0.114
Protein	0.122	0.208	0.007	0.188	0.016
Riboflavin	0.124	0.126	0.109	0.168	0.101
Saturated fat	0.125	0.158	0.090	0.070	0.111
Sodium	0.025	0.104	0.047	0.121	0.046
Total fat	0.120	0.181	0.062	0.034	0.087
Thiamin	0.074	0.201	0.113	0.235	-0.015
Vitamin B6	0.145	0.181	0.129	0.199	0.063
Vitamin B12	0.036	0.100	0.108	0.167	0.122
Vitamin A	0.034	0.032	0.078	0.041	0.069
Vitamin C	0.124	0.032	0.171	0.106	0.062
Vitamin E	0.074	0.115	0.148	0.094	0.048
Water	0.167	0.130	0.185	0.243	0.130
Zinc	0.095	0.243	0.049	0.146	-0.001

Table 2. (Continued)

Component	Males 20-59	Males ≥ 60	Females 20-59	Females ≥ 60	Persons 0-19
Percent tfat	0.085	0.066	0.062	-0.004	0.030
Percent mfat	0.110	0.094	0.034	-0.041	0.002
Percent sfat	0.134	0.107	0.119	0.037	0.128
Percent pfat	0.056	0.086	0.077	0.031	0.031
Percent carb	0.118	0.071	0.097	-0.021	0.052
Percent prot	0.078	0.059	0.115	0.107	-0.010

The model estimates for niacin, protein, and zinc for individuals less than 20 years of age were -0.011, -0.004, and -0.006, respectively. These negative estimates are not significantly different from zero. Similarly, model estimates for the correlations for calories from total fat, monounsaturated fat, and carbohydrates, for women 60 years and older, were also negative though not significantly different from zero. Because only a few direct estimates in the entire data set are negative

Table 3. Correlations among estimated day-to-day correlations

	Males 20-59	Males ≥ 60	Females 20-59	Females ≥ 60	Persons 0-19
Males 21-59	1.00	0.34	0.35	0.17	0.32
Males ≥ 60	0.34	1.00	0.08	0.57	0.17
Females 21-59	0.35	0.08	1.00	0.49	0.50
Females ≥ 60	0.16	0.57	0.49	1.00	0.31
Persons 0-20	0.32	0.17	0.50	0.31	1.00

Table 4. Estimated day-to-day correlations based upon measurement error model

Component	Males 20-59	Males ≥ 60	Females 20-59	Females ≥ 60	Persons 0-19
Calcium	0.113	0.137	0.128	0.139	0.119
Carbohydrates	0.129	0.162	0.165	0.205	0.164
Carotene	0.089	0.121	0.076	0.082	0.045
Cholesterol	0.076	0.048	0.039	-0.070	0.027
Copper	0.112	0.166	0.130	0.193	0.106
Fiber	0.111	0.196	0.131	0.247	0.092
Folate	0.106	0.124	0.112	0.107	0.100
Iron	0.086	0.169	0.076	0.168	0.020
Energy	0.115	0.158	0.136	0.180	0.120
Magnesium	0.123	0.210	0.159	0.287	0.129
Monounsaturated fat	0.093	0.091	0.079	0.031	0.066
Niacin	0.078	0.183	0.061	0.185	-0.011
Polyunsaturated fat	0.088	0.106	0.070	0.053	0.045
Phosphorus	0.103	0.194	0.115	0.235	0.068
Potassium	0.119	0.189	0.147	0.244	0.122
Protein	0.080	0.179	0.063	0.180	-0.004
Riboflavin	0.108	0.150	0.120	0.158	0.100
Saturated fat	0.106	0.109	0.109	0.079	0.103
Sodium	0.085	0.133	0.068	0.100	0.026
Total fat	0.096	0.100	0.088	0.051	0.076
Thiamin	0.084	0.211	0.077	0.245	0.001
Vitamin B6	0.104	0.179	0.115	0.208	0.077
Vitamin B12	0.109	0.142	0.120	0.143	0.104
Vitamin A	0.091	0.087	0.075	0.020	0.062
Vitamin C	0.105	0.122	0.110	0.103	0.098
Vitamin E	0.098	0.130	0.095	0.108	0.070
Water	0.127	0.182	0.163	0.239	0.150
Zinc	0.079	0.172	0.060	0.165	-0.006

Table 4. (Continued)

Component	Males 20-59	Males ≥ 60	Females 20-59	Females ≥ 60	Persons 0-19
Percent tfat	0.082	0.079	0.055	-0.005	0.035
Percent mfat	0.073	0.071	0.035	-0.031	0.007
Percent sfat	0.112	0.087	0.119	0.045	0.131
Percent pfat	0.084	0.097	0.061	0.031	0.034
Percent carb	0.091	0.071	0.074	-0.010	0.068
Percent prot	0.082	0.136	0.063	0.103	0.017

and because there is no behavioral reason to expect negative day-to-day correlations, we suggest that estimates of zero be used for these component groups. The smooth estimate for the day-to-day correlation for cholesterol intake in women 60 years of age was negative, and significantly different from zero. Examination of the data showed that this correlation estimate was primarily due to two individuals who had cholesterol intakes that were outliers when compared to the rest of the individuals. Therefore, we suggest that this estimate also be set to zero.

The estimation computations are described in Appendix A. The vector of standard errors for the estimates in a row of Table 4 is (0.0064, 0.0201, 0.0139, 0.0385, 0.0223). Thus, the use of the model produces estimates with variances much smaller than the variances of the original estimates. The variance of the model estimates for males 20-59 is approximately one fifth of the original variance while the model variances for males older than 59 and females 20-59 are less than 50 percent of the variance of the direct estimates.

Table 5 contains average correlations for females greater than 19 years of age and males greater than 19 years of age. The entries in Table 5 were obtained from those in Table 4 by weighting the two entries for the two age groups in Table 4 by the fraction of respondents in each of the two groups.

Summary and Conclusions

The estimated correlation coefficients presented in Tables 2, 4, and 5 were obtained from the combined CSFII data for the years, 1989, 1990, and 1991. Direct estimates of the day-to-day correlation coefficients are shown in Table 2. Estimated correlation coefficients are positively correlated. See Table 3. In addition, the variability of mean correlations (over all dietary components) for each group is small relative to estimation error. Therefore, smooth estimates of the day-to-day correlations were obtained by fitting a measurement error model to the direct correlation estimates. See Tables 4 and 5. These smoothed correlation estimates have considerably smaller variances than the direct correlation estimates.

The estimated correlations can be used as input parameters in the ISU software, to obtain estimates of usual intake distributions of various dietary components. Because the smoothed correlation estimates presented in Table 4 have uniformly smaller variance than the direct estimates. It is suggested that the smooth estimates be used in the ISU procedure.

Correlation coefficients were obtained from five subpopulations: males 20-59, males 60 and older, females 20-59, females 60 and older, and individuals under 20 years of age. While dietary assessments are performed for subpopulations that are further disaggregated by race, other age groups and income, it is suggested that aggregated correlation estimates be used. Correlation coefficients estimated from subpopulations with few individuals have high standard errors and may lead to unreliable estimates of the within and the between individual variances computed by the ISU method.

Most of the correlation coefficients that were computed are significantly different from zero. This was to be expected, since for all age-gender groups, sample sizes were large. Those computed for energy, protein, vitamins A and C, iron and calcium, are in agreement with the estimates obtained from the 1987 National Food Consumption Survey (An and Carriquiry 1992).

Table 5. Estimated day-to-day correlations based upon measurement error model

Component	Males ≥ 20	Females ≥ 20	Persons 0-19
Calcium	0.119	0.131	0.119
Carbohydrates	0.138	0.178	0.164
Carotene	0.097	0.078	0.045
Cholesterol	0.069	0.005	0.027
Copper	0.126	0.150	0.106
Fiber	0.133	0.167	0.092
Folate	0.111	0.110	0.100
Iron	0.108	0.105	0.020
Energy	0.126	0.150	0.120
Magnesium	0.146	0.199	0.129
Monounsaturated fat	0.092	0.064	0.066
Niacin	0.106	0.100	-0.011
Polyunsaturated fat	0.093	0.065	0.045
Phosphorus	0.127	0.153	0.068
Potassium	0.137	0.178	0.122
Protein	0.106	0.100	-0.004
Riboflavin	0.119	0.132	0.100
Saturated fat	0.107	0.100	0.103
Sodium	0.098	0.078	0.026
Total fat	0.097	0.076	0.076
Thiamin	0.117	0.130	0.001
Vitamin B6	0.124	0.144	0.077
Vitamin B12	0.118	0.127	0.104
Vitamin A	0.090	0.058	0.062
Vitamin C	0.109	0.108	0.098
Vitamin E	0.106	0.099	0.070
Water	0.141	0.187	0.150
Zinc	0.103	0.093	-0.006

Table 5. (Continued)

Component	Males ≥ 20	Females ≥ 20	Persons 0-19
Percent tfat	0.081	0.036	0.035
Percent mfat	0.072	0.014	0.007
Percent sfat	0.105	0.096	0.131
Percent pfat	0.087	0.052	0.034
Percent carb	0.086	0.048	0.068
Percent prot	0.096	0.076	0.017

APPENDIX A. STATISTICS FOR CALCULATING ESTIMATES

In this appendix, we present some of the statistics used in the calculation of the estimates of Table 3. Our approach follows the methodology described in Fuller (1987, Section 4.1). In our analysis, we treat $\sum_{ee}$ as known.

The estimated covariance matrix of $\hat{\rho}_k = (\hat{\rho}_{k1}, \hat{\rho}_{k2}, \hat{\rho}_{k3}, \hat{\rho}_{k4}, \hat{\rho}_{k5})$ is

$$\mathbf{m}_{YY} = \begin{pmatrix} 1.211 & 0.784 & 0.565 & 0.556 & 0.558 \\ 0.784 & 4.318 & 0.253 & 3.455 & 0.571 \\ 0.565 & 0.253 & 2.099 & 2.073 & 1.156 \\ 0.556 & 3.455 & 2.073 & 8.618 & 1.458 \\ 0.558 & 0.571 & 8.618 & 1.458 & 2.554 \end{pmatrix} \times 10^{-3},$$

where

$$\begin{aligned} \mathbf{m}_{YY} &= (34)^{-1} \sum_{k=1}^{33} (\hat{\rho}_k - \bar{\rho}_k)' (\hat{\rho}_k - \bar{\rho}_k), \\ \bar{\rho}_k &= (34)^{-1} \sum_{k=1}^{34} \hat{\rho}_k. \end{aligned}$$

The five roots of the characteristic equation

$$|\mathbf{m}_{YY} - \lambda \Sigma_{ee}| = 0$$

are 7.47, 3.19, 1.59, 1.21, and 0.45. The test statistic given in (4.1.48) of Fuller (1987, p. 303) is

$$F = 33\,[3\,(31)]^{-1}\,(1.59 + 1.24 + 0.45) = 1.16.$$

If the two-factor model is true, the statistic is approximately distributed as a central F with 93 and ∞ degrees of freedom. Therefore, by this criterion, the data are consistent with the model.

The three characteristic vectors of the matrix Σ_{ee} associated with the three smallest roots are the columns of the matrix

$$\hat{\mathbf{B}} = \begin{pmatrix} 2.551 & 25.612 & -16.731 \\ 9.073 & 7.208 & 12.447 \\ -26.706 & 7.210 & 13.443 \\ -0.214 & -6.090 & -7.719 \\ 16.342 & -10.202 & -0.912 \end{pmatrix},$$

where we use $\hat{\mathbf{B}}$ for the matrix to agree with the notation of Fuller (1987, p. 293). The estimate of the β-matrix is

$$\begin{bmatrix} \hat{\beta}_{11} & \hat{\beta}_{12} & \hat{\beta}_{13} \\ \hat{\beta}_{21} & \hat{\beta}_{22} & \hat{\beta}_{23} \end{bmatrix} = \begin{bmatrix} 0.486 & 0.504 & 0.168 \\ 0.284 & -0.173 & 0.580 \end{bmatrix},$$

with estimated covariance matrix,

$$\hat{\Sigma}_{\beta} = \begin{bmatrix} 7.574 & -7.538 & 0.082 & -0.081 & 0.870 & -0.866 \\ -7.538 & 30.106 & -0.081 & 0.325 & -0.870 & 3.458 \\ 0.082 & -0.081 & 24.401 & -24.287 & 0.603 & -0.600 \\ -0.081 & 0.325 & -24.287 & 96.996 & -0.600 & 2.400 \\ 0.870 & -0.870 & 0.603 & -0.600 & 7.560 & -7.525 \\ -0.866 & 3.458 & -0.600 & 2.400 & -7.525 & 30.052 \end{bmatrix} \times 10^{-3}.$$

The estimated covariance matrix is computed using Theorem 4.1.5 of Fuller (1987, p. 305). The estimate of the mean square of the true ρ_{k4} and ρ_{k5} for females 60 years and older, and persons less than 20 years of age are

$$\hat{M}_{kk44} = 0.00690, \qquad \hat{M}_{kk55} = 0.00181.$$

The estimates of the true values in the ℓ-th row of Table 4 is

$$\hat{\mathbf{r}}_{.\ell} = \bar{\rho} + (\hat{\rho} - \bar{\rho})\left(\mathbf{I} - \hat{\mathbf{B}}\hat{\mathbf{B}}'\Sigma_{ee}\right),$$

where $\hat{\mathrm{r}}_{.\ell}$ is the ℓ-th row of the table, $\bar{\rho}$ is the vector of mean direct correlation estimates over all dietary components, water, and ratios of dietary components, and $\hat{\rho}_{.\ell}$ is the vector of original estimated correlations for the ℓ-th component. The estimated covariance matrix of the errors in the estimated day-to-day correlations of Table 4 is

$$\hat{V}\left\{\hat{\mathrm{r}}_{.\ell} - \mathrm{r}'_{.\ell}\right\} = \mathbf{\Sigma}_{ee} - \mathbf{\Sigma}_{ee}\hat{\mathrm{B}}\hat{\mathrm{B}}'\mathbf{\Sigma}_{ee}.$$

The diagonal elements of this matrix are

$$(0.042,\ 0.405,\ 0.195,\ 1.483,\ 0.500) \times 10^{-3}.$$

REFERENCES

An, A. and A. L. Carriquiry. 1992. Estimation of the correlation among days for transformed nutrient intake. Report prepared for the Human Nutrition Information Service. October, 1992. Iowa State University, Ames.

Battese, G. E., S. M. Nusser, and W. A. Fuller. 1988. Estimation of the distribution of usual intakes for selected dietary components using data from the 1985-86 CSFII. Draft report submitted to USDA, Human Nutrition Information Service. April. Iowa State University.

Carriquiry, A. L., W. A. Fuller, and J. J. Goyeneche. 1995. Estimation of usual intake distribution of ratios. Preliminary report prepared for USDA, Agricultural Research Service. April. Iowa State University.

Fuller, W. A. 1987. Measurement Error Models. Wiley, New York.

Lin, L. I-K, and E. F. Vonesh. 1989. An empirical nonlinear data-fitting approach for transforming data to normality. *The American Statistician* 43, 237-243.

National Academy of Sciences, Food Nutrition Board. 1986. Nutrient Adequacy: Assessment Using Food Consumption Surveys. National Academy Press, Washington, D.C.

Neter, J., Wasserman, W., and M. H. Kutner. 1990. Applied Linear Statistical Models. 3rd ed. Irwin, Homewood, Illinois.

Nusser, S. M., A. L. Carriquiry, K. W. Dodd, and W. A. Fuller. 1995. A semiparametric approach to estimating usual daily intake distributions. Tentatively accepted by the *Journal of the American Statistical Association.*