Method of Moments Estimation of Usual Nutrient Intake Distributions

by S. M. Nusser,
G. E. Battese, and W. A. Fuller

Working Paper 90-WP 52
March 1990

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

S. M. Nusser is a graduate assistant with CARD and the ISU Department of Statistics. G. E. Battese is senior lecturer, Department of Econometrics, University of New England, Armidale, N.S.W., Australia. W. A. Fuller is distinguished professor, Department of Statistics, Iowa State University.

This research was supported by the Human Nutrition Information Service, USDA, Cooperative Agreement No. 58-3198-6-60 and No. 58-3198-9-032.

ABSTRACT

The parameters of the distribution of usual intakes of calcium, energy, iron, protein and vitamin C are estimated using four daily intakes of these dietary components for a sample of women selected in the Continuing Survey of Food Intakes by Individuals, conducted in 1985-86 by the Human Nutrition Information Service of the U.S. Department of Agriculture. The daily intakes for an individual are assumed to be the sum of the usual intake for that individual plus a measurement error. The variance and the third moment of the measurement errors for a given individual are assumed to be functions of the usual intake of that individual. For the five dietary components investigated, the standard deviation and cube root of the third moment of the measurement errors of an individual both appear to be linearly related to the usual intake for that individual. Moments of usual intake are estimated as functions of the sample moments of the four-day mean intakes and estimators for the parameters of the distribution of the measurement errors. Given estimates of the moments of usual intakes, the parameters of a particular distribution function can be estimated. The assumption that usual intakes of dietary components are distributed as Weibull random variables and that the measurement errors are generated as deviations of observations from the mean of Weibull distributions was accepted for all dietary components.

METHOD OF MOMENTS ESTIMATION OF
USUAL NUTRIENT INTAKE DISTRIBUTIONS

1. INTRODUCTION

The United States Department of Agriculture has been responsible for conducting periodic surveys to estimate food consumption patterns of households and individuals in the United States since 1936. Data from these surveys have had a significant impact on the formulation of food-assistance programs, on consumer education, and on food regulatory activities.

In evaluating the adequacy of diets it is recognized that an individual who has a low intake of a given dietary component on one day is not necessarily deficient (or at risk of being deficient) so far as that dietary component is concerned. It is low intake over a sufficiently long period of time that produces a dietary deficiency. A dietary deficiency exists when the usual (i.e., normal or long-run

George E. Battese is Senior Lecturer, Department of Econometrics, University of New England, Armidale, N.S.W. 2351, Australia. Sarah M. Nusser and Wayne A. Fuller are Graduate Assistant and Distinguished Professor, respectively, Department of Statistics, Iowa State University, Ames, Iowa 50011. This research was partly supported by Research Agreement No. 58-3198-6-60, between the Human Nutrition Information Service, U.S. Department of Agriculture and the Center for Agricultural and Rural Development, Iowa State University. The authors thank Stanley Johnson and Helen Jensen for comments on earlier drafts and Carol Meeter for computer programming assistance. A part of this research was conducted when George Battese was on a study leave from the University of New England.

average) intake of the dietary component is less than the appropriate dietary standard. The same concepts apply to excessive intakes.

To assess usual intake, daily dietary intakes are often collected from individuals for a number of days. An individual's mean daily intake of a particular dietary component is then used as an indication of the individual's usual intake for that component. While mean intake is an improved measure of usual intake for an individual, the distribution of mean intakes does not adequately represent the distribution of usual intakes. In general, the variance of the mean intakes will be greater than the variance of usual intakes. Discrepancies for other distributional parameters, such as skewness, may also exist. Because of these problems, using the mean intake distribution as an estimate of the usual intake distribution can lead to erroneous inferences regarding nutritional status. For example, if the mean intake distribution is used to estimate the proportion of the population whose usual intakes fall below an intake level indicative of dietary deficiency, the estimated proportion will be overestimated due to the overdispersion of the mean intake distribution relative to the usual intake distribution.

Two alternative approaches to estimating the usual intake distribution are: (1) transforming the observed intakes to normality and (2) modeling the data in the original scale. We focused our first research efforts on modeling the intakes in the original scale. We adopted an approach to estimating the distribution of usual intakes that relies on a measurement error model for observed intakes and on models relating the intra-individual measurement error variance and third moment to an individual's true usual intake. On the basis of these models, method of moments estimators for usual intake moments were developed. The estimated usual intake moments were then used to estimate the parameters of the distribution of usual intakes.

2. DIETARY INTAKE DATA

We based our analyses on data from the Continuing Survey of Food Intakes by Individuals, conducted by the United States Department of Agriculture in 1985-86. Daily dietary intakes were collected from women between 19 and 50 years of age and from the pre-school children of the women. Daily intakes were to be obtained at approximate two-month intervals over the period of one year (April 1985 to March 1986). Data for the first day were collected by personal interview and were based on a 24-hour recall. Data for subsequent days were based on 24-hour recall and were collected by telephone whenever possible. The sample was a multi-stage stratified area probability sample from the 48 coterminous states. The primary sampling units were area segments, and the probabilities of selection of area segments were proportional to the numbers of housing units in the segments as estimated by the Bureau of the Census. Because of the high rate of nonresponse for the six-day sample, the United States Department of Agriculture constructed a four-day data set for analyses. The four days of data consisted of the first day of dietary intakes for all individuals who provided at least four days of data, plus a random selection of three daily intakes from the remaining three, four or five days of data available.

In this paper we analyze a subset of the four-day data set containing dietary intakes for women between 23 and 50 years of age who were responsible for meal planning within the household and who were not pregnant or lactating during the survey period. There were 785 women who belonged to this category. Empirical results are presented for intakes of five dietary components: calcium, energy, iron, protein and vitamin C. These components were selected because of their nutritional importance, and because they include diverse intake behaviors and metabolic and storage properties.

3. PRELIMINARY ANALYSES

The report of the National Research Council (1986) provides a comprehensive review of factors that influence observed daily intake data. The effect of some of these variables, such as errors in reported food intake and translation of food intake to nutrient intake, are not estimable from the data available for our study. The effect of other factors, such as day of the week, season (month), interview method and interview sequence can be investigated with this data.

The daily intake data were examined using analysis of variance methods to determine whether weekday, month, interview method (personal or telephone) and interview sequence effects were important. Interview sequence refers to the order in which the daily data were obtained for sample individuals; there were four values corresponding to this variable. Preliminary analyses with weekday, month, interview method and interview sequence effects in the model indicated that month and interview sequence are confounded to a large degree. This is because the first interview was conducted at nearly the same point in time for all individuals. Hence, the month effects were deleted from the model, and for subsequent analyses, a model involving weekday, interview method and interview sequence as additive classification variables was used.

Interview method was not found to be significant for any dietary component. Weekday effects were significant for energy ($p < 0.001$) and protein ($p < 0.01$) intakes. Contrasts indicated that the effect was primarily due to higher consumption on weekends for both dietary components. Weekday effects were not significant for calcium, iron or vitamin C. Sequence effects (confounded with month effects) were significant at the $\alpha = 0.001$ level for calcium, energy, iron and protein intakes. For all dietary components, a large proportion of the sequence variation was accounted for by a contrast of first interview day versus the intake for the other three days (92-99% of the sequence variation for calcium, energy, iron and protein; 78% for vitamin C). The mean intakes for the first interviews, conducted April through June, were consistently higher than mean intakes in other months.

Because of these results, we used data adjusted for weekday and interview sequence effects in the subsequent analyses. Instead of using the usual linear adjustment (residuals from the analysis of variance), we used a ratio adjustment to insure that all adjusted intake values were nonnegative. To implement the ratio adjustment, the observed intake values were regressed on class variables representing the days of the week and the interview sequence. Predicted values were calculated and the data were adjusted using

$$Y_{ij} = \hat{Y}^{o\,-1}_{ij}\,\overline{Y}^{o}_{..}\,Y^{o}_{ij} \; ,$$

where Y^{o}_{ij} is the original observed intake of individual i on day j , $\overline{Y}^{o}_{..}$ is the grand mean of the original observed intakes, $\hat{Y}^{o}_{ij}$

Table 1. Analysis of variance for observed individual intakes

Source	d.f.	S.S.[a]	EMS
Individuals	784	$\sum_{i=1}^{n} 4(\bar{Y}_{i.} - \bar{Y}_{..})^2$	$\sigma_w^2 + 4\sigma_a^2$
Days/individual	2355	$\sum_{i=1}^{n}\sum_{j=1}^{4} (Y_{ij} - \bar{Y}_{i.})^2$	σ_w^2

[a] $\bar{Y}_{i.}$ is the average intake over four days for individual i; $\bar{Y}_{..}$ is the average intake over all observations on the 785 individuals.

is the predicted intake from the regression, Y_{ij} is the ratio adjusted intake, and $i=1, 2, \ldots, n=785$ individuals and $j=1, 2, \ldots, r=4$ days. All of the analyses discussed in this paper are performed on the ratio adjusted observed intakes, which are hereafter called the observed intakes.

To investigate the level of variation among individuals (inter-individual variance) relative to the variation among days for a given individual (intra-individual variance), the simple analysis of variance of Table 1 was constructed. The inter-individual variance, σ_a^2, and the intra-individual variance, σ_w^2, were estimated from this analysis of variance table. The square roots of the estimates for the intra-individual variance are given in the first column of Table 2; the ratios of the estimated intra-individual variance to the inter-individual variance are presented in the second column of Table 2. These estimates indicate that the intra-individual variance is about twice as large as the inter-individual variance of the daily intakes for the five dietary components involved. The ratios observed in this study are similar to the ratios reported by Sempos et al. (1985), given in the last column of Table 2.

Basic features of the distributions of four-day average intakes of the five dietary components were obtained by calculating estimates for the mean, standard deviation, skewness and kurtosis (SAS, 1985, pp.737-741). Given the k-th central moment of the means, denoted by $\mu_{k\bar{Y}}$, skewness is $\mu_{3\bar{Y}} / \mu_{2\bar{Y}}^{3/2}$, and kurtosis is $\mu_{4\bar{Y}} / \mu_{2\bar{Y}}^{2} - 3$. The estimates are presented in Table 3. The skewness estimates indicate that the distributions of the four-day average intakes are skewed to the right for all five dietary components. The energy and protein intakes are the least skewed. The kurtosis values indicate that the distributions of the four-day average intakes for the five dietary components tend to have fatter tails than the normal distribution.

Table 2. Estimates for the intra-individual standard deviation and the ratio of intra-individual to inter-individual variances of daily intakes for each dietary component

Dietary component	$\hat{\sigma}_w$	$\hat{\sigma}_w^2/\hat{\sigma}_a^2$ This study	$\hat{\sigma}_w^2/\hat{\sigma}_a^2$ Sempos et al. (1985)
Calcium	310.3	1.8	1.1[a]
Energy	546.0	1.8	1.6
Iron	4.6	2.5	2.6
Protein	25.4	2.8	2.1
Vitamin C	60.7	2.4	2.4

[a]Average of two ratio estimates for adult women in two different years.

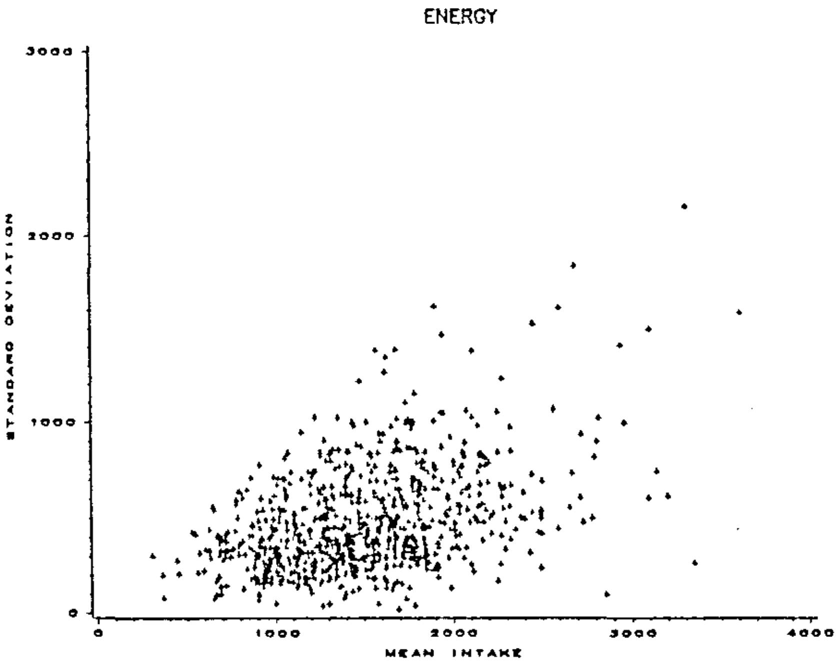
Table 3. Summary statistics for the distribution of the four-day average intakes for each dietary component

Dietary component	Mean	Standard deviation	Skewness	Kurtosis
Calcium (10 mg)	57.90	28.14	1.16	2.40
Energy (100 kcal)	14.93	4.91	0.65	0.95
Iron (mg)	9.99	3.70	1.23	3.50
Protein (g)	59.56	19.75	0.75	2.22
Vitamin C (mg)	75.15	49.84	1.39	2.84

To study the conditional distribution of daily intakes for an individual, we considered the sample variances of the four daily intakes for each sample individual. Let the sample variance of daily intakes for the i-th individual be denoted by

$$S_i^2 = \frac{1}{r-1} \sum_{j=1}^{r} (Y_{ij} - \bar{Y}_{i.})^2 , \qquad (1)$$

where $r = 4$ is the number of observations per individual. Plots of the sample standard deviation, S_i , against the average daily intake, $\bar{Y}_{i.}$, for the sample individuals are presented in Figure 1 for energy and vitamin C. These two dietary components represent extremes for many distributional characteristics. It is evident that the sample standard deviations of the individual intakes tend to increase as the four-day averages increase. These plots, and plots not shown for the other dietary components, suggest that the true standard deviation of individual intakes may be directly proportional to the usual intake of the individual.



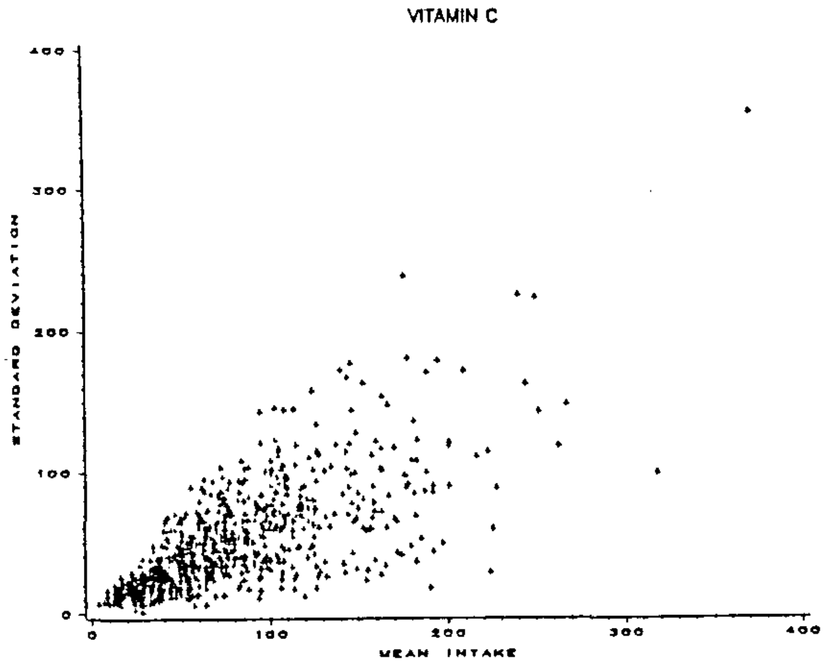


Figure 1. Plot of the sample standard deviations of daily intakes against the average intakes of women 23-50 years old for energy and vitamin C.

The extent to which the intra-individual third moment is related to the mean daily intakes for individuals is also of interest. The third moment about the mean of the daily intakes for individuals is unbiasedly estimated by

$$M_{3i} = \frac{r}{(r-1)(r-2)} \sum_{j=1}^{r} (Y_{ij} - \bar{Y}_{i.})^3 \tag{2}$$

for a sample of r days. Figure 2 presents the plot of the cube root of M_{3i} against the average daily intake, $\bar{Y}_{i.}$, for energy and vitamin C. These plots, and similar ones for the other dietary components, suggest that the cube root of the third moment of the individual daily intakes may be a linear function of the usual intake. A more detailed discussion of models for the moments of individual daily intakes is given in the next section.

4. DISTRIBUTION OF USUAL INTAKES

The concept of the usual intake of a dietary component for an individual is crucial to our study. The usual intake of a dietary component for the i-th individual is defined to be the conditional expectation of the daily intakes of that dietary component for individual i, and is denoted by y_i; i.e.,

$$y_i = E(Y_{ij} | i) .$$

One can think of the usual intake for an individual as the average of daily intakes where the average is over a sufficiently long period of time.

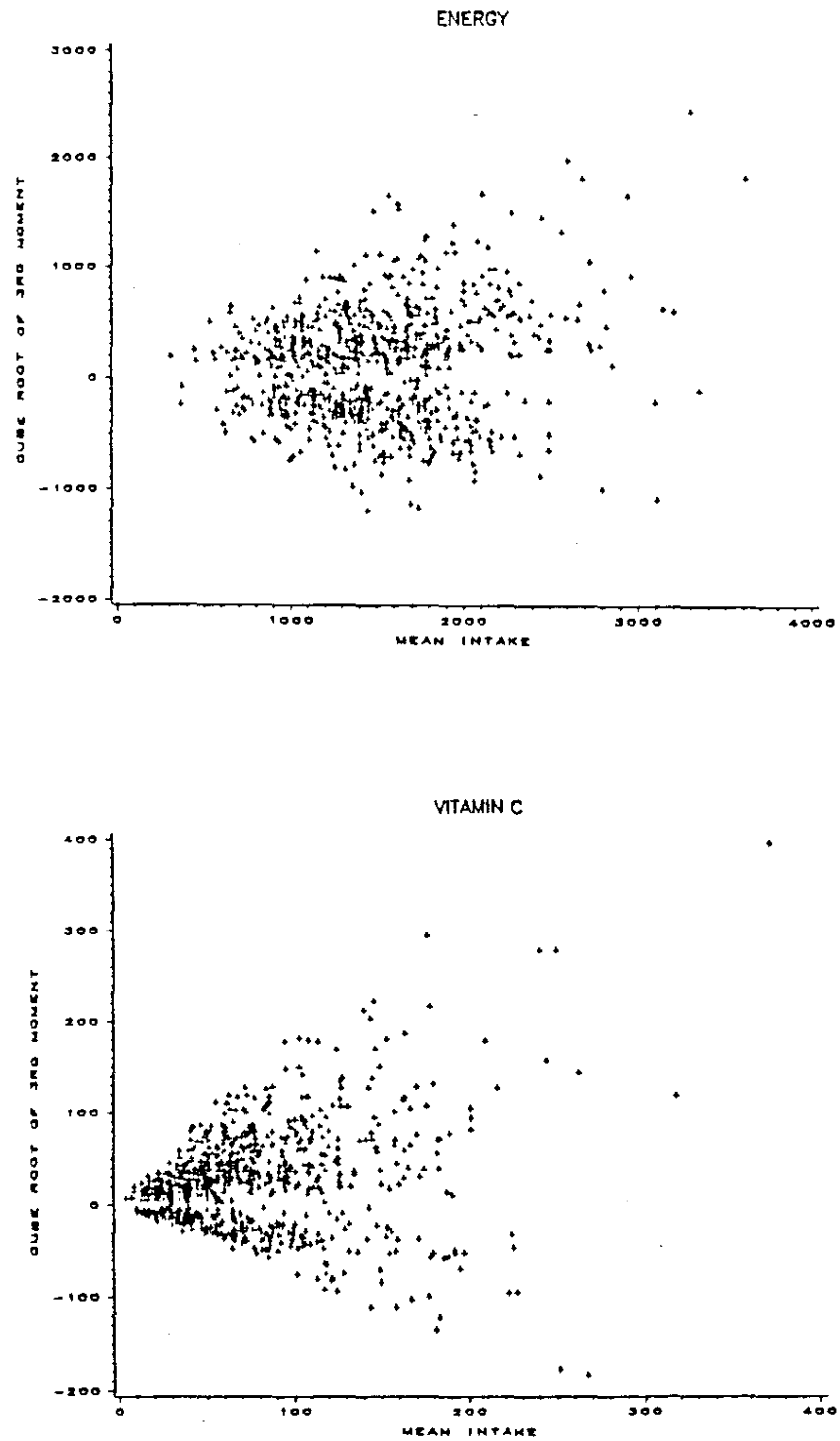


Figure 2. Plot of the cube roots of the sample third moment of daily intakes against the average intakes of women 23-50 years old for energy and vitamin C.

We begin by considering a measurement error model in which the daily intakes of dietary components are expressed as the sum of usual intake and a measurement error, and where the moments of the measurement error have a particular structure. This model is described and estimators of the parameters are presented in the next section.

4.1. Model for Measurement Errors

Suppose that a random sample of n individuals from the population is available and that r daily intakes for a dietary component are available for each individual. The difference between the reported daily intake for the i-th individual on the j-th reporting day and the usual daily intake for the i-th individual, $Y_{ij} - y_i = e_{ij}$, is called the measurement error associated with reported daily intakes. Under the definition of usual intake, the measurement errors, e_{ij}, $j=1, 2, \ldots, r$, have zero mean for all individuals, $i=1, 2, \ldots, n$.

Our model is

$$Y_{ij} = y_i + e_{ij}, \quad j=1, 2, \ldots, r; \quad i=1, 2, \ldots, n, \tag{3}$$

where

$$E(e_{ij}^2 | i) = \alpha\, y_i^2, \quad i=1, 2, \ldots, n, \tag{4}$$

$$E(e_{ij}^3 | i) = \gamma\, y_i^3, \quad i=1, 2, \ldots, n, \tag{5}$$

and the sixth moments of the measurement errors exist. We also assume that the measurement errors for the i-th individual, $e_{i1}, e_{i2}, \ldots, e_{ir}$, are (conditionally) independent and that the measurement errors for different individuals, e_{ij} and $e_{k\ell}$, where $i \neq k$, are independent.

Under the model specification (3)-(5), the standard deviations of the measurement errors and the cube roots of the third moments of the measurement errors are directly proportional to the usual intakes. Models (4) and (5) are consistent with the plots of Figures 1 and 2, respectively.

Estimators for the parameters, α and γ, are derived as follows. It is easily verified that unbiased estimators for $E(e_{ij}^2 | i)$ and $E(e_{ij}^3 | i)$ are S_i^2, defined in (1), and M_{3i}, defined in (2), respectively. Also, unbiased estimators for y_i^2 and y_i^3 are $[\bar{Y}_{i.}^2 - r^{-1}S_i^2]$ and $[\bar{Y}_{i.}^3 - 3r^{-1}\bar{Y}_{i.} S_i^2 + 2r^{-2}M_{3i}]$, respectively. Thus estimators for α and γ are

$$\hat{\alpha} = \frac{\sum_{i=1}^{n} S_i^2}{\sum_{i=1}^{n} (\bar{Y}_{i.}^2 - r^{-1}S_i^2)} \tag{6}$$

and

$$\hat{\gamma} = \frac{\sum_{i=1}^{n} M_{3i}}{\sum_{i=1}^{n} [\bar{Y}_{i.}^3 - 3r^{-1}\bar{Y}_{i.} S_i^2 + 2r^{-2}M_{3i}]} . \tag{7}$$

Values of the estimators and their estimated standard errors, obtained using standard results for ratio estimators, are presented in Table 4.

4.2. Moments of the Usual Intakes

We assume that the set of usual intakes of a dietary component for n individuals, $y_1, y_2, \ldots, y_n$, is a random sample from a distribution with finite fourth moment. The population mean of the

Table 4. Estimated parameters (and estimated standard errors) of the measurement error models for each dietary component

	Dietary Component				
Parameter	Calcium	Energy	Iron	Protein	Vitamin C
α	0.2479 (0.0133)	0.1240 (0.0054)	0.1929 (0.0118)	0.1710 (0.0078)	0.5169 (0.0294)
γ	0.1298 (0.0258)	0.0353 (0.0070)	0.1543 (0.0328)	0.0630 (0.0112)	0.5220 (0.112)

usual intakes is represented by μ_y . The second, third and fourth central moments of the usual intakes are denoted by μ_{ky} , $k=2, 3, 4$, respectively, where

$$\mu_{ky} \equiv E(y_i - \mu_y)^k .$$

The moments of the average of four daily intakes for a random sample of individuals can be expressed in terms of the moments of the usual intakes and the parameters of the measurement error model (4)-(5). The estimators for the first four moments are presented in the Appendix.

The variance for the estimator for the mean of the usual intakes, $\hat{\mu}_y = \bar{Y}_{..}$, can be unbiasedly estimated with

$$\hat{V}ar(\hat{\mu}_y) = [n(n-1)]^{-1} \sum_{i=1}^{n} (\bar{Y}_{i.} - \bar{Y}_{..})^2 .$$

Approximate standard errors for the remaining estimated parameters were generated using a jackknife method. The full data set was randomly partitioned into twenty groups. Twenty data sets were then generated by omitting one of the twenty partitions for each data set. The parameters were estimated for each of the twenty data sets. An estimate of the variance for each parameter estimator was calculated as

$$\hat{V}ar(\hat{\eta}) = K^{-1}(K-1) \sum_{k=1}^{K} (\hat{\eta}_k - \bar{\eta}_.)^2$$

where $\hat{\eta}_k$ is the estimate of η derived from the k-th data set (k=1, 2, ..., K=20) , and $\bar{\eta}_. = K^{-1}\Sigma_{k=1}^{K}\hat{\eta}_k$.

Estimates for the first four moments of the usual intakes are presented in Table 5 in the form of roots of the estimators. In addition, estimates for the skewness and kurtosis parameters, β_1 and β_2 , are given. Estimated standard errors for the estimators for these parameters, computed by the jackknife method, are given in parentheses below the estimates.

There are several differences between the statistics of Table 3 and those of Table 5. The estimated variances of the usual intakes range from 70 percent of the estimated variance of the four-day mean intake for calcium to 59 percent of the estimated variance of the four-day mean intake for protein. Thus the variability of daily intakes makes an important contribution to the total variability of the four-day means. In all cases, the estimated skewness of the distribution of the usual intakes is less than that for the four-day means, although the distributions of the usual intakes remain positively skewed. The distribution of usual intakes for energy and protein appear to be more symmetric than the usual intake distributions for calcium, iron and vitamin C. The estimated kurtosis of usual intakes is also smaller than

Table 5. Estimates of the moments (and estimated standard errors) of the distribution of usual intakes for dietary components

Dietary Component	Parameters					
	$\hat{\mu}_y$	$\hat{\mu}_{2y}^{1/2}$	$\hat{\mu}_{3y}^{1/3}$	$\hat{\mu}_{4y}^{1/4}$	Skewness $\hat{\beta}_1$	Kurtosis $\hat{\beta}_2$
Calcium (10 mg)	57.90 (0.10)	23.45 (1.06)	21.63 (3.21)	33.29 (3.25)	0.76 (0.27)	1.06 (1.10)
Energy (100 kcal)	14.93 (0.18)	4.09 (0.17)	2.69 (0.60)	5.56 (0.27)	0.29 (0.16)	0.44 (0.45)
Iron (mg)	9.99 (0.13)	2.91 (0.20)	2.82 (0.49)	4.50 (0.49)	0.91 (0.32)	2.67 (1.33)
Protein (g)	59.56 (0.70)	15.11 (0.95)	10.50 (7.6)	23.53 (1.9)	0.33 (0.42)	2.88 (0.90)
Vitamin C (mg)	75.15 (1.8)	39.41 (1.0)	34.11 (5.14)	51.60 (5.53)	0.65 (0.23)	-0.06 (0.93)

the estimated kurtosis of four-day means for all dietary components except protein. The kurtosis of the usual intakes of energy and vitamin C differ little from zero, the kurtosis of the normal distribution.

4.3. Weibull Model

While the moments of usual intake are informative, it is the distribution function that is of real interest. In this section we develop parametric models for the distribution of usual intake and for the distribution of the measurement errors. Our basic model remains that of (3)-(5). We assume that the distribution of usual intakes, is a member of the three-parameter Weibull distribution with density function

$$f(y) = \theta^{-\beta}\, \beta(y - \nu)^{\beta-1} \exp(-[\theta^{-1}(y - \nu)]^{\beta}) \ , \ y > \nu \ , \ \theta > 0, \ \beta > 0 \ , \ \nu > 0 \ ,$$

where θ is the scale parameter, β is the shape parameter, and ν is the shift parameter which determines the minimum value of the distribution. If $\beta > 1$, the density function is unimodal and positively skewed with a value of zero at $y = \nu$.

The mean of the Weibull distribution is

$$E(y_i) = \nu + \theta\, \Gamma(1 + \beta^{-1}) \,, \tag{8}$$

where $\Gamma(\ )$ is the gamma function, and the second and third moments about the mean are

$$E[y_i - E(y_i)]^2 = \theta^2[\Gamma(1 + 2\beta^{-1}) - \Gamma^2(1 + \beta^{-1})] \tag{9}$$

$$E[y_i - E(y_i)]^3 = \theta^3[\Gamma(1 + 3\beta^{-1}) - 3\Gamma(1 + 2\beta^{-1})\, \Gamma(1 + \beta^{-1}) + 2\Gamma^3(1 + \beta^{-1})] \,, \tag{10}$$

respectively.

For each dietary component, the shift parameter ν was set at 20% of the RDA as defined in National Research Council (1980). This value was chosen becuse it seemed physiologically sound to require a positive minimum usual intake value. In addition, preliminary results for the case $\nu = 0$ indicated that the estimated third moment of usual intakes, $\hat{\mu}_{3y}$, was not well fit by the simpler two-parameter Weibull distribution. Third moments under the three-parameter Weibull model with $\nu = .20$ RDA were much closer to the estimated $\hat{\mu}_{3y}$ for each dietary component.

Method-of-moments estimators for the scale and shape parameters of the distribution of the usual intakes were obtained by equating the right hand side of (8) and (9) to $\hat{\mu}_y$ and $\hat{\mu}_{2y}$, the estimators for the mean and variance of the usual intakes. We used the IMSL routine DNEQNF, a Levenburg-Marquardt algorithm with a finite-difference approximation to the Jacobian, to solve the nonlinear system for $\hat{\theta}$ and $\hat{\beta}$. Estimates of the parameters and their jackknife standard errors for calcium, energy, iron, protein and vitamin C are presented in Table 6.

To complete our model for daily intakes, we assume that the measurement errors for the i-th individual, $e_{i1}, e_{i2}, \ldots, e_{ir}$, are (conditionally) independent random variables defined by

$$e_{ij} = Z_{ij} - E(Z_{ij}|i) \ , \quad j=1, 2, \ldots, r \ ,$$

where $\{Z_{i1}, Z_{i2}, \ldots, Z_{ir}\}$ is a random sample from a two-parameter Weibull distribution with scale and shape parameters, θ_{ei} and β_e .

Table 6. Scale and shape estimates (and estimated standard errors) for the Weibull distribution of usual intakes

Dietary component	RDA	ν	$\hat{\theta}$	$\hat{\beta}$
Calcium (mg)	800	160.0	471.78 (9.97)	1.854 (0.095)
Energy (kcal)	2000[a]	400.0	1225.89 (19.76)	2.908 (0.120)
Iron (mg)	18	3.6	7.22 (0.16)	2.331 (0.165)
Protein (g)	44	8.8	56.21 (1.08)	3.744 (0.233)
Vitamin C (mg)	60	12.0	70.60 (1.84)	1.645 (0.067)

[a]The value for energy is the mean energy requirement.

These assumptions imply that the Weibull distributions associated with the measurement errors on different individuals have different scale parameters, θ_{ei} , $i=1, 2, \ldots, n$, but a common shape parameter, β_e . Given that the model is constrained to satisfy the moment properties defined by equations (4) and (5), it follows that $\theta_{ei} = \delta y_i$, $i=1, 2, \ldots, n$, where δ is a positive constant. The equations defining the method-of-moments estimators of δ and β_e are

$$\hat{\alpha} = \hat{\delta}^2[\Gamma(1 + 2\hat{\beta}_e^{-1}) - \Gamma^2(1 + \hat{\beta}_e^{-1})]$$

$$\hat{\gamma} = \hat{\delta}^3[\Gamma(1 + 3\hat{\beta}_e^{-1}) - 3\Gamma(1 + 2\hat{\beta}_e^{-1})\ \Gamma(1 + \hat{\beta}_e^{-1}) + 2\Gamma^3(1 + 3\hat{\beta}_e^{-1})] \ ,$$

where $\hat{\alpha}$ and $\hat{\gamma}$ are defined by (6) and (7). Values of the estimators for δ and β_e for the five dietary components and their jackknife standard errors are given in Table 7.

To test the fit of the hypothesized Weibull distributions, Monte Carlo methods were used to generate the distribution of individual four-day mean intakes from the estimated Weibull distributions for the usual intakes and the corresponding measurement errors. For each dietary component, 100,000 usual intakes, y_i , were generated along with four measurement errors, e_{ij} , $j=1, 2, 3, 4$, for each y_i according to the parameters of the respective estimated Weibull distributions. The usual intake plus the corresponding average measurement error, $y_i + \bar{e}_{i.}$, were used to generate a cumulative distribution function against which the empirical distribution function for the observed four-day mean intakes could be compared. The hypothesized cumulative distribution function for individual four-day means was generated by

Table 7. Parameter estimates (and estimated standard errors) for the Weibull distribution of the measurement errors

Dietary component	$\hat{\delta}$	$\hat{\beta}_e$
Calcium	0.822 (0.054)	1.52 (0.12)
Energy	0.675 (0.063)	1.76 (0.18)
Iron	0.479 (0.049)	1.06 (0.10)
Protein	0.753 (0.070)	1.67 (0.17)
Vitamin C	0.972 (0.086)	1.26 (0.12)

counting the number of generated observations contained in each of 1000 intervals over the range of the observed four-day mean intakes. A chi-square goodness-of-fit statistic, involving thirty mutually exclusive intervals over the range of the four-day mean intakes, was used as the test statistic. Observed frequencies for the generated four-day mean intakes and the observed four-day mean intakes were tabulated in a 2×30 table, and expected values were generated. Values of the test statistic, based on 27 degrees of freedom, are listed in Table 8 for each of the five dietary components. Tests of size 0.05 indicate that assuming the usual intakes have a Weibull distribution and the measurement errors are generated from Weibull distributions is satisfactory for all dietary components except calcium. However, the value of the test statistic for calcium is relatively close to the critical value, indicating that the departure is not severe. Plots comparing the empirical cumulative distribution function of the four-day

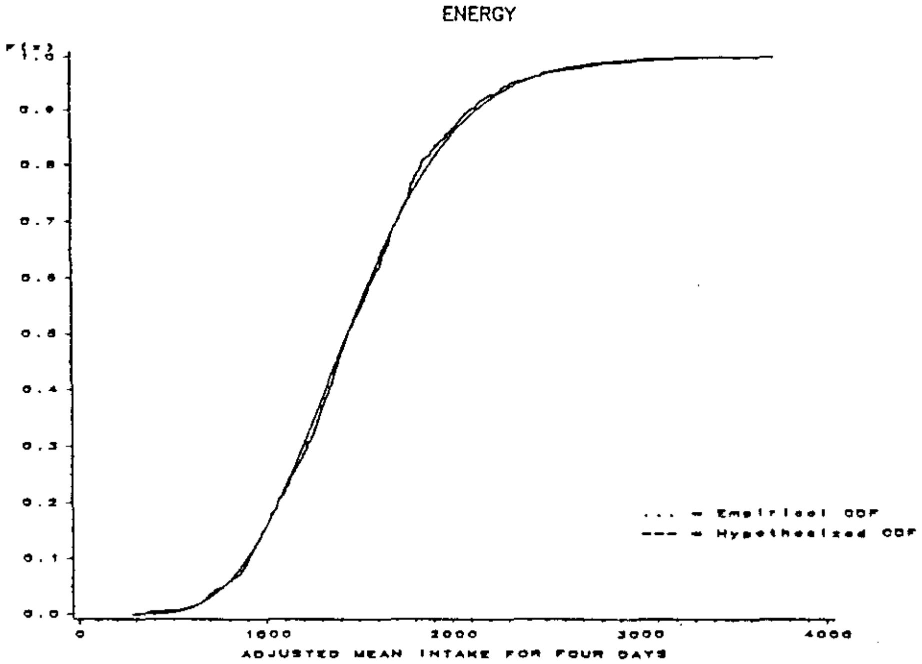
Table 8. Goodness-of-fit statistics for the distribution of four-day mean intakes based on Weibull distributions

Dietary component	χ^2 [a]
Calcium	43.7
Energy	30.8
Iron	39.5
Protein	22.9
Vitamin C	32.4

[a]The 95-th percentile for the chi-square distribution with 27 degrees of freedom is 40.1.

mean intakes with the hypothesized cumulative distribution function based on Weibull densities for energy and vitamin C are shown in Figure 3.

To illustrate the use of the estimated distribution functions consider vitamin C. The recommended daily allowance of vitamin C for non-pregnant, non-lactating, 23-50 year-old women is 60 mg (National Research Council, 1980). On the basis of the estimated Weibull distribution, 22% of women have usual intakes less than 42 mg, 70% of the recommended daily allowance. If the distribution of means is used as an estimate of the usual intake distribution, an estimated 30% of the women have vitamin C usual intakes less than 42 mg. Note that the overdispersion in the mean distribution relative to the usual intake distribution leads to an overestimate of the percentage of women whose usual intakes fall below 70% of the recommended daily allowance. This illustrates the importance of obtaining accurate estimates of usual intake distributions for use in developing policies to improve the nutritional status of the population.



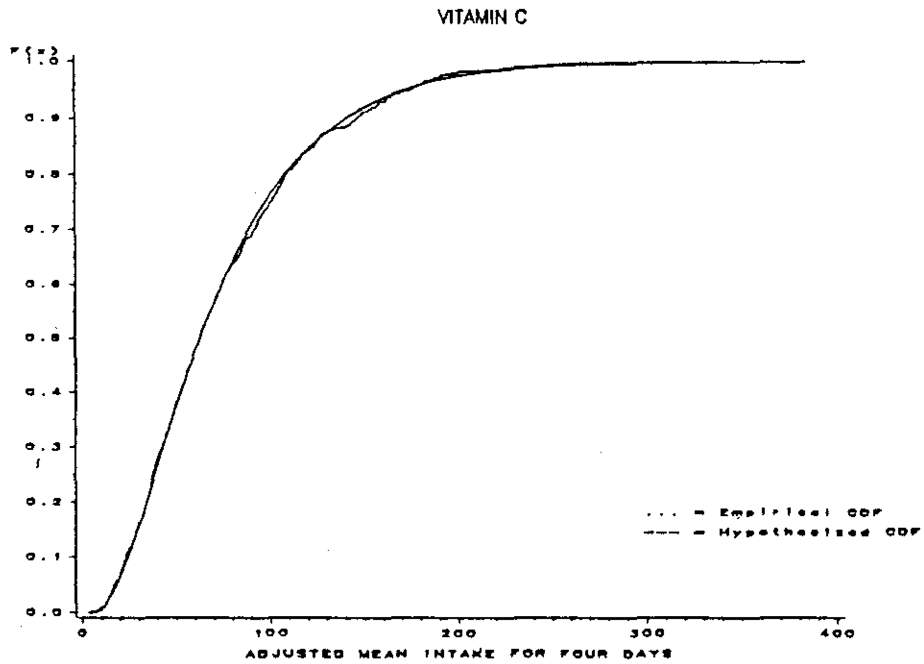


Figure 3. Plots comparing the empirical cdf of the four-day mean intakes with the hypothesized cdf based on Weibull densities for energy and vitamin C.

APPENDIX

Moments of the Usual Intakes

We express the first four moments of the usual intakes, y_i, in terms of moments of $\bar{Y}_{i.} = y_i + \bar{e}_{i.}$ and the parameters, α and γ, of the model (4)-(5). To do this, consider the expected value of deviations of mean individual intake from the grand mean to the second, third and fourth power, summed over individuals. Given the assumptions of the model (3)-(5) it can be shown that

$$E\{\sum_{i=1}^{n} (\bar{Y}_{i.} - \bar{Y}_{..})^2\} = (n-1)\mu_{2\bar{Y}} ,$$

$$E\{\sum_{i=1}^{n} (\bar{Y}_{i.} - \bar{Y}_{..})^3\} = n^{-1}(n-1)(n-2)\mu_{3\bar{Y}} ,$$

and

$$E\{\sum_{i=1}^{n} (\bar{Y}_{i.} - \bar{Y}_{..})^4\} = n^{-2}(n-1)\{(n^2 - 3n + 3)\mu_{4\bar{Y}} + 3(2n - 3)(\mu_{2\bar{Y}})^2\} ,$$

where $\mu_{k\bar{Y}} = E(\bar{Y}_{i.} - \mu_{\bar{Y}})^k$, $k=2, 3, 4$.

Using the relation $\bar{Y}_{i.} - \mu_{\bar{Y}} = (y_i - \mu_y) + \bar{e}_{i.}$ and the error moment models, it can be shown that

$$\mu_{2\bar{Y}} = [1 + r^{-1}\alpha]\mu_{2y} + (r^{-1})\alpha\, \mu_y^2$$

$$\mu_{3\bar{Y}} = [1 + 3(r^{-1})\alpha + (r^{-2})\gamma]\mu_{3y} + 3(r^{-2})\mu_y[2r\alpha + \gamma]\mu_{2y} + (r^{-2})\gamma\mu_y^3$$

$$\mu_{4\bar{Y}} = \{1 + 6(r^{-1})\alpha + 4(r^{-2})\gamma + 3(r^{-3})(r-1-c)\alpha^2\}\mu_{4y}$$
$$+ \{12(r^{-3})\mu_y[r^2\alpha + r\gamma + (r-1-c)\alpha^2]\mu_{3y}$$
$$+ \{6(r^{-3})\mu_y^2[r^2\alpha + 2r\gamma + 3(r-1-c)\alpha^2]\}\mu_{2y} + 3(r^{-3})(r-1-c)\alpha^2\mu_y^4$$
$$+ [r(r-1)(r^2 - 3r + 3)]^{-1}E[\sum_{j=1}^{r}(Y_{ij} - \bar{Y}_{i.})^4] .$$

The above expressions for $\mu_{\bar{Y}}$, $\mu_{2\bar{Y}}$, $\mu_{3\bar{Y}}$, and $\mu_{4\bar{Y}}$, can be used to derive method-of-moment estimators for the first four moments of the usual intakes, μ_y , μ_{2y} , μ_{3y} and μ_{4y} . These estimators are

$$\hat{\mu}_y = \bar{Y}_{..} ,$$

$$\hat{\mu}_{2y} = \{\hat{\mu}_{2\bar{Y}} - r^{-1}\hat{\alpha}\,\hat{\mu}_y^2\}[1 + r^{-1}\hat{\alpha}]^{-1} ,$$

$$\hat{\mu}_{3y} = \{\hat{\mu}_{3\bar{Y}} - 3(r^{-2})\hat{\mu}_y(2r\hat{\alpha} + \hat{\gamma})\hat{\mu}_{2y} - (r^{-2})\hat{\gamma}\,\hat{\mu}_y^3\}[1 + 3(r^{-1})\hat{\alpha} + (r^{-2})\hat{\gamma}]^{-1} ,$$

$$\hat{\mu}_{4y} = \{\hat{\mu}_{4\bar{Y}} - [12(r^{-3})\hat{\mu}_y(r^2\hat{\alpha} + r\hat{\gamma} + (r-1-c)\hat{\alpha}^2)]\hat{\mu}_{3y}$$
$$- [6(r^{-3})\hat{\mu}_y^2(r^2\hat{\alpha} + 2r\hat{\gamma} + 3(r-1-c)\hat{\alpha}^2)]\hat{\mu}_{2y} - 3(r^{-3})(r-1-c)\hat{\alpha}^2\hat{\mu}_y^4$$
$$- \frac{1}{n}\sum_{i=1}^{n} M_{4i}\}[1 + 6(r^{-1})\hat{\alpha} + 4(r^{-2})\hat{\gamma} + 3(r^{-3})(r-1-c)\hat{\alpha}^2]^{-1} ,$$

where

$$\hat{\mu}_{2\bar{Y}} = \frac{1}{n-1}\sum_{i=1}^{n}(\bar{Y}_{i.} - \bar{Y}_{..})^2 ,$$

$$\hat{\mu}_{3\bar{Y}} = \frac{n}{(n-1)(n-2)} \sum_{i=1}^{n} (\bar{Y}_{i.} - \bar{Y}_{..})^3$$

$$\hat{\mu}_{4\bar{Y}} = \{\frac{n^2}{n-1} \sum_{i=1}^{n} (\bar{Y}_{i.} - \bar{Y}_{..})^4 - 3(2n - 3)(\hat{\mu}_{2\bar{Y}})^2\}(n^2 - 3n + 3)^{-1} ,$$

and

$$M_{4i} = \frac{1}{r(r-1)(r^2 - 3r + 3)} \sum_{j=1}^{r} (Y_{ij} - \bar{Y}_{i.})^4 .$$

REFERENCES

National Research Council (1980), Recommended Daily Allowances, National Academy Press, Washington, D.C.

National Research Council (1986), Nutrient Adequacy: Assessment Using Food Consumption Surveys, National Academy Press, Washington, D.C.

SAS Institute Inc. (1985), SAS User's Guide: Basics, 1985 Edition, Cary, North Carolina.

Sempos, C. T., Johnson, N. E., Smith, E. L. and Gilligan, C. (1985), "Effects of intraindividual and interindividual variation in repeated dietary records," American Journal of Epidemiology **121**, 120-130.