Fruit and vegetable consumption in Malaysia: a count system approach

Steven T. Yen
Department of Agricultural and Resource Economics
The University of Tennessee
Knoxville, TN 37996-4518, USA
E-mail: syen@utk.edu

Andrew K.G. Tan
School of Social Sciences
Universiti Sains Malaysia
11800 Penang, Malaysia
E-mail: atan@usm.my

Paper prepared for presentation at the EAAE 2011 Congress
Change and Uncertainty
Challenges for Agriculture,
Food and Natural Resources

August 30 to September 2, 2011
ETH Zurich, Zurich, Switzerland

1. Introduction

The keys to good health and disease prevention are exercise and good dietary habits. Fruits and vegetables are (FV) specifically known to be important for health because they are naturally low in calories and provide essential nutrients and dietary fiber. They may also play a role in preventing certain chronic diseases. When compared to people who eat only small amounts of fruits and vegetables, those who eat more generous amounts, as part of a healthy diet, tend to have reduced risk of chronic diseases. These diseases include stroke, type 2 diabetes, some types of cancer, and perhaps cardiovascular disease and hypertension (USDHHS and USDA, 2005). The World Health Organization (WHO) recommends a daily diet of five servings of fruits and vegetables (FV) to prevent heart diseases, cancer, diabetes, obesity and other diseases (WHO, 2003). It is estimated that low FV consumption causes about 19% of gastrointestinal cancers, 31% of ischemic heart diseases, and 11% of strokes worldwide in 2002. Low FV intake is also ranked as one of the top 10 global mortality risk factors, and up to 2.7 million lives could potentially be saved annually given sufficient FV consumption (WHO, 2003). In spite of these well known facts, few people consume enough FV, opting instead for processed foods which have a more consistent taste and a longer shelf life.

Statistics from the Food and Agriculture Organization (FAO) of the United Nations indicate that between 1980 and 2003, FV consumption in Malaysia averaged about 150 grams of fruits and 78 grams of vegetables per capita per day (FAOSTAT, 2009). However, the combined FV consumption of 228 grams per day was far below the 400 grams, or five servings, recommended by WHO dietary guidelines. This suggests that Malaysians are not consuming enough of FV compared to their consumption of other staple foods such as meat and rice.

Given the well known health benefits of FV consumption, it is important to identify the socio-demographic determinants of FV consumption and the profile of people who do not eat enough FV. Various studies have investigated this issue in western countries (Cox and Wohlgenant, 1986; Huang, 1993; You et al., 1998; Feng and Chern, 2000; Huang and Lin, 2000). Nevertheless, no known study has examined the socio-demographic determinants of FV consumption in Malaysia. As a result, this study attempts to augment this research gap by providing insights into the influence of socio-demographic characteristics on the demand for FV. Further, despite the WHO guidelines on combined FV consumption, we draw on the empirical literature of typically separate analyses of FV and investigate consumption of the two products separately. In general, the outcomes of the study are important to public health policy makers concerned with the nutritional status of the population and to FV industry analysts interested in distinguishing their target market.

This study differs from previous FV consumption studies in two important respects. First, a bivariate count equation system is developed to accommodate both the discrete nature of and correlation between the dependent variables. Second, unlike previous studies which mainly focused on developed western countries, we focus on a newly industrialized country, Malaysia, for which few demand studies have existed.

2. Methodology

Count data models have been used in modeling consumer food demand (e.g., Lee, 1985), and we extend the single equation models to a bivariate count equation system framework to account for the discrete nature of and correlation between the two dependent variables of FV, measured in terms of days of servings per week. The general approach to our bivariate specification is to begin with a univariate count cumulative distribution function (CDF) and then link the two probabilities with a copula. In what follows, observation subscripts are suppressed for brevity.

We consider two count distributions for each dependent variable y_j. The first is the Poisson distribution, which has CDFs

$$F_j(y_j) = \sum_{h=0}^{y_j} e^{-\mu_j} \mu_j^h / h! \text{ for } j = 1,2, \tag{1}$$

with the conditional mean of y_j parameterized as

$$\mu_j = \exp(x'\beta_j). \tag{2}$$

The Poisson distribution is known to be restrictive due to its equidispersion property, *viz*., with the mean and variance both equal to μ_j. This restriction can be unpalatable in most empirical applications. To accommodate overdispersion in the data, we considered several and settle with one form of the negative binomial distribution, known as NB1 (Cameron and Trivedi, 1986), with CDFs

$$F_j(y_j) = \sum_{h=0}^{y_j} \frac{\Gamma(\mu_j/\alpha_j + h)}{\Gamma(\mu_j/\alpha_j)\Gamma(h+1)} \left(\frac{1}{1+\alpha_j}\right)^{\mu_j/\alpha_j} \left(\frac{\alpha_j}{1+\alpha_j}\right)^h, \alpha_j \geq 0,\ j = 1,2, \tag{3}$$

in which the conditional mean μ_j is also parameterized as in equation (2), and the conditional variance as

$$\omega_j = \mu_j + \alpha_j \mu_j, \tag{4}$$

where α_j is the overdispersion parameter, and $\Gamma(\zeta) = \int_0^\infty t^{\zeta-1} e^{-t} dt$ is the Gamma function of ζ for $\zeta > 0$.[1] Thus, overdispersion is admitted when $\alpha_j > 0$, and the NB1 reduces to the Poisson distribution when $\alpha_j = 0$, a testable parametric restriction.

Although the FV demand equations can be estimated as separate count-data models using the Poisson distribution (1) or negative binomial distribution (3), a major shortcoming with such single-equation models is that correlation between the count variables is not accommodated. In the context of count dependent variables, such correlation may lie in unobserved heterogeneity and joint estimation taking account of correlated errors produces more efficient estimates (Cameron and Trivedi, 2005, p. 685). Existing multivariate count models, often in the form of a bivariate system, are restrictive. The moment-based approach of Gouriéroux et al. (1984), for instance, does not maintain the integer-valued property of the counts—a major shortcoming. Other bivariate Poisson or negative binomial models are restrictive in that the error correlation is restricted to be positive (Kocherlakota and Kocherlakota, 1993; Marshall and Olkin, 1990). Procedures to develop less restrictive bivariate count distributions tend to be computationally complex (Cameron and Trivedi, 1998). To overcome such shortcomings in bivariate count-data modeling, we follow the copula approach, which has the advantage of accommodating both

---

1 Cameron and Trivedi (1986) propose a general negative binomial distribution with a variance function $\mu + \alpha\mu^p$, where p is an estimable parameter. The NB1 is a special case of that general distribution when $p = 1$. The NB2, with $p = 2$, was rejected by the fruit data (with convergence failure), and estimation of this general model produced an estimate of p that was not significantly different from 1 for fruits, supporting NB1. All forms of negative binomial distribution (NB1, NB2 and with a free parameter p) were rejected by the vegetable data so the margin for vegetables is specified as Poisson.

positive and negative error correlations (see Cameron et al., 2004 for the copula approach to other multivariate count-data models). The two count CDFs, of any legitimate (flexible) forms, are linked with a copula function to form a bivariate count distribution.

We present only the Gaussian copula, identified as the preferred copula in the current study after a model specification search. The Gaussian copula is characterized by a two-place function of random variables F_1 and F_2 (which are marginal CDFs in the current case)

$$C(F_1, F_2; \rho) = \Phi_2[\Phi^{-1}(F_1), \Phi^{-1}(F_2); \rho], \tag{5}$$

where Φ and Φ_2 are the univariate and bivariate standard normal CDFs, respectively, and ρ is Pearson's correlation coefficient between the random variables defining the margins F_1 and F_2. This is the distribution function used by Lee (1983), not called copula at the time, in developing sample selection models with non-Gaussian error distributions. For maximum-likelihood (ML) estimation, we need the copula density. By taking finite differences, the copula density is (Cameron et al., 2004)

$$\begin{aligned} c[F_1(y_1), F_2(y_2); \rho] = C[F_1(y_1), F_2(y_2); \rho] - C[F_1(y_1-1), F_2(y_2); \rho] \\ - C[F_1(y_1), F_2(y_2-1); \rho] + C[F_1(y_1-1), F_2(y_2-1); \rho]. \end{aligned} \tag{6}$$

This density is the likelihood contribution for an individual observation, and the sample likelihood function for an independence sample is the product of the densities for all sample observations. To recapitulate the copula approach, specification of the count-data system consists of two steps: (a) choice of a copula such as the Gaussian copula in (5), and (b) specifications of CDFs such as Poisson in (1) or NB1 in (2). Upon estimation, marginal effects of explanatory variables can be derived by differentiating the conditional mean (2).

3. Data and variable definitions

Data for this study were obtained from the Malaysia Non-Communicable Disease Surveillance-1 (MyNCDS-1) by the Ministry of Health Malaysia (2006). The nationwide survey encompassed the thirteen states and the federal territory of Kuala Lumpur. The survey period lasted from September 2005 to February 2006 and was conducted based on a two-stage stratified random sampling procedure to ensure the representativeness of the sample of the Malaysian population. Inclusion criteria are Malaysian citizens between 25–64 years of age, across genders, and ethnic groups. While a total of 3040 individuals responded to the survey, 2572 respondents were retained, with a total response rate of 84.6%. The final sample in the current study contains 2447 observations after excluding those observations with missing or suspect data. Further details on the survey and data collection are available in the survey documentation (Ministry of Health Malaysia, 2006).

The dependent variables, denoted by FV consumption, were collected as counts: days consumed per week. Due to the lack of empirical research on the subject in Malaysia, the selection of variables hypothesized to influence FV demand relies on previous studies by Blisard et al. (2004), Stewart et al. (2004), Gustavsen and Rickertsen (2002, 2006), and Casagrande et al. (2007). Based on insights from these works, the following socio-demographic variables are posited to influence FV consumer demand: (i) length of typical work day, (ii) education levels, (iii) age brackets, (iv) ethnicity/race, (v) income levels, (vi) gender, (vii) marital status, (viii) location of residence, (ix) smoking status, (x) health status, and (xi) regional location (refer to Table 1 for variable definitions and sample statistics).

Specifically, education level is represented by the highest level of formal education possessed by the respondent (Primary, Junior high, Senior high and Tertiary). Age brackets are denoted by younger (Age ≤ 30), middle-age younger (Age 31–40), middle-age older (Age 41–58), and retiree (Age ≥ 59), with the expectation that differences in age and life-cycle patterns may lead to changes in FV preferences and consumption patterns. The unique racial composition in Malaysia, consisting of the three primary ethnic groups of Malay, Chinese, and Indian, along with those of other ethnicities (base group), allow for an ideal setting to examine the role of ethnicity in determining FV consumption patterns. Monthly household income was originally collected in 10 categorical intervals, but we re-defined it in terms of five categories: poverty, low, middle-low, middle-high and high income groups (see Table 1 for coding). Gender is entered as a dummy variable (Male) with a value of 1 for males and 0 for females. To examine the importance of access to food stores, location of residence (Rural) is denoted by 1 if the individual reside in a rural area and 0 otherwise.

Several unique socio-demographic variables which are hypothesized to have an important impact on FV consumption are considered in the present study. Since smokers are deemed to have lower concerns for their health and well-being compared to non-smokers, it is expected that an inverse relationship will result between smoking status and the demand for FV. Thus, smokers are designated by a value of 1 and 0 otherwise. Additionally, current health status is conjectured to positively affect FV demand as individuals diagnosed with hypercholesterolemia (Hyperchol), hypertension (High BP), and diabetes (Diabetes) are assigned a value of 1 for the presence of each condition and 0 otherwise.

Since data from MyNCDS-1 do not contain information on price, we utilize regional variables as proxies, with the assumption that the standard of living (and hence price) in metropolitan areas would be higher compared to those in less-metropolitan surroundings. Respondents are categorized into those from Region 1 (consisting of the metropolitan states of Penang, Selangor and the Federal Territory in Peninsular Malaysia), Region 2 (the less-metropolitan Peninsular Malaysia states of Perlis, Kedah, Perak, Melaka, Negeri Sembilan, Johor, Pahang, Kelantan, and Terengganu in Peninsular Malaysia) as the reference group, and Region 3 (the East Malaysian states of Sabah and Sarawak). Besides price variations, these regional variables might also reflect other sources of regional differences. Finally, the length of a typical work day (Work hours) is used as a proxy for the time available for healthy food consumption.

4. Maximum-likelihood estimation and model selection

The system of FV equations is estimated after a model search using a non-nested specification test procedure to select a copula. Specifically, let r_i and s_i be the maximum log likelihood contributions of sample observation i for two competing specifications and define differences $d_i = r_i - s_i$ for $i = 1,\dots, n$, with sample mean $\overline{d}$ and standard deviation s_d. Then, under the null hypothesis of no difference between the two models, Vuong's (1989, equations (3.1), (4.2), (5.6)) standard normal statistic is $z = n^{1/2}\overline{d} / s_d \sim \mathcal{N}(0,1)$.

Three different copulas (Gaussian, Frank, and Clayton) are considered, along with four different margins for each of the two equations (Poisson, NB1, NB2, negative binomial with a free parameter; see Footnote 1).[2] Regardless of the copula chosen, all three forms of the negative

2 The Frank and Clayton copulas are not presented due to space consideration. See Nelsen (2006, p. 116) for details on these copulas.

binomial distribution are rejected by the vegetable data, and estimation of the general negative binomial system produces an estimate for the free parameter (p) not significantly different from 1 for fruits, suggesting NB1. Further, Vuong's non-nested tests suggest use of the Gaussian copula over the Frank (z = 2.67, p-value = 0.0075) and Clayton (z = 3.19, p-value = 0.0014) copulas. Therefore, the count system is estimated with the Gaussian copula, with NB1 margin for fruits and Poisson margin for vegetables. The ML estimate for the overdispersion parameter (α_1) is significantly different from 0 at the 1% level of significance, suggests overdispersion of days per week for fruits, justifying the NB1 distribution for fruits over the Poisson distribution (Table 2). Further, estimate of the error correlation (ρ) is significantly different from 0 at the 1% level of significance, which justifies the system estimation to improve statistical efficiency.

As the effects of explanatory variables can be examined in greater depth by calculating marginal effects, we defer further discussions of the empirical results to the next section.

5. Marginal effects of explanatory variables

The marginal effects of explanatory variables on the mean days of FV servings per week, calculated by differentiating the conditional mean (2), are presented in Table 3. Education levels play an important role in promoting consumption of fruits as evidenced by the significant and positive relationship between education levels and days of consumption per week. Specifically, compared to individuals with only primary school education, those with junior high school (tertiary) education consume 0.27 (0.63) more day of fruits per week, while senior high school education does not have a statistically significant impact.

FV consumption increases as one becomes older. Those in middle- (retiree-) age group consume 0.69 (0.90) more day of fruits per week than those in the younger age groups. In comparison to consumers of other ethnicities, Malay (0.50), Chinese (0.68) and Indian (0,65) respondents consume more days of fruits per week. In contrast, individuals in the poverty (1.10), low (0.79), and middle-low (0,33) income groups consume less days of fruit per week than those in the high income group. Smokers eat 0.32 less day of fruits per week than non-smokers. Respondents with hypercholesterolemia consume 0.23 more day of fruits per week than their cohort without high cholesterol levels, holding other factors constant. Residents of metropolitan areas (Region 1) eat 0.33 less day, while those in the East Malaysian states of Sabah and Sarawak (Region 3) consume 0.77 more day, of fruits per week than those in the less-metropolitan Peninsular Malaysia states (Region 2).

Income also plays a role in determining vegetable consumption. This is evidenced by the fewer days of vegetables servings per week by those in the poverty (0.85) and low (0.76) income groups compared to high income earners. A likely explanation could be that lower income consumers do not consider vegetables as a necessity in their diets, while favoring other more filling food products such as rice and meats. Last, metropolitan residents in Peninsular Malaysia consume 1.21 fewer days of vegetables per week compared to non-metropolitan residents in Peninsular Malaysia, *ceteris paribus*. These results can be rationalized by the notion that the higher prices or other lifestyle factors may be a factor for metropolitan consumers to consume less FV compared to their non-metropolitan cohorts.

6. Discussion and concluding remarks

Some of the world's most widespread and debilitating nutritional disorders, including birth defects, mental and physical retardation, weakened immune systems, blindness, and even death, are caused by diets lacking in FV. Diets high in FV are also associated with a reduced risk for chronic diseases. In addition, since FV have low energy density (i.e., few calories relative to volume), eating them as part of a reduced-calorie diet can be beneficial for weight management. Encouraging people to eat more FV is therefore often at the top of the agenda for nutrition educators. Still, most populations are not consuming nearly enough FV, including Malaysia.

While a number of studies on FV consumption have focused on western countries, scant information is available on the determinants of FV consumption in developing countries. In this paper, we analyze the determinants of FV consumption in Malaysia with a bivariate count data model using the copula approach. Such models developed by the copula approach have the advantage of being flexible in that overdispersion is allowed and error correlation is not restricted, and the estimation procedure is simple. Results of the study indicate that socio-demographic factors such as education, age, ethnicity, income, smoking status, region and health conditions significantly affect FV consumption.

Specifically, education levels positively affect consumption of fruits as those with at least high school education consume significantly more in terms of days per week compared to those with only primary school education. Individuals between 41–58 years of age and those 59 years and above consume more FV than those below 30 years of age. Individuals of Malay, Chinese and Indian ethnicities eat more fruits than those of other ethnic backgrounds. Findings also suggest that low income individuals tend to have fewer days per week of FV consumed compared to higher income individuals.

The responses of smokers toward fruit consumption are less encouraging compared to non-smokers. In addition, health conditions of individuals are significant contributors of FV consumption as evidenced by the higher consumption rates of fruits amongst those diagnosed with hypercholesterolemia compared to those without this ailment. It is also interesting to note that consumers in the metropolitan areas consume less FV compared to those in the less metropolitan areas. Plausible explanations may be that prices or even lifestyle differences between these localities cause consumers to have differing FV consumption rates. Interestingly, contrary to expectations that employment might take time away from preparation of FV for servings, work hours do not affect FV consumption.

Our results imply a need for measures to educate and motivate consumers to make healthier dietary choices. Interventions that increase FV consumption by changing behaviors should be considered, as should those that increase public awareness of the benefits of FV in the diet. However, nutritional interventions should go beyond increasing awareness and targeting groups of individuals. These programs should attempt to eliminate barriers to healthy eating, provide support for persons who are making healthy changes, increase resources for populations with greater need, and emphasize nutritional policies that have an impact on the society. Simply put, these intervention programs should be targeted at and tailored toward those who have lower FV consumption. Based on our findings, these groups in Malaysia generally include the less educated, the young, the poor, smokers, and those living in metropolitan surroundings.

As the virtues of consuming FV are often taught in schools, it may be worthwhile to continue educating younger individuals about the benefits of FV, especially given our findings that they are less likely to consume FV than older individuals. Hence, educational programs should target the less educated and younger age groups in order to sustain lifelong beneficial effects. Also, since poorer individuals consume less FV than middle and high income individuals, government policies toward providing food assistance to the poor could be geared specifically toward increasing FV consumption.

While this study provides interesting new findings for a developing country like Malaysia, similar analysis should also be conducted in other countries to assess the robustness of our findings. Intervention programs that are specifically geared toward increasing FV consumption in targeted groups should be evaluated to assess their effectiveness and applicability in different settings. Future research might focus on identification of barriers to eating more FV and on evaluating environmental changes that could potentially increase FV consumption (e.g., increasing the proportion of FV in vending machines; promoting healthful food advertising and availability of healthful foods).

Last, while the lack of price and quantity/expenditure data in health surveys may be quite common, it certainly is a limitation of the present study. Future studies should replicate our analysis using consumption or expenditure data and also with longitudinal panel data to assess the robustness of our findings. Other commonly considered variables in such demand studies may also include the presence of children in the household, family size and occupation type.

References

Blisard, N., Stewart, H., Jolliffe, D., 2004. Low-income households' expenditures on fruits and vegetables. Agricultural Economic Report No. 833, U.S. Department of Agriculture, Economic Research Service, Washington DC, May, 30 pp.

Cameron, A.C., Trivedi, P.K., 1986. Econometric models based on count data: comparisons and implications of some estimators and tests. Journal of Applied Econometrics 1(1), 29–53.

Cameron, A.C., Trivedi, P.K., 1998. Regression Analysis of Count Data. Cambridge University Press, New York.

Cameron, A.C., Trivedi, P.K., 2005. Microeconometrics: Methods and Applications. Cambridge University Press, New York.

Cameron, A. C., Tong, L., Trivedi, P.K., Zimmer, D.M., 2004. Modelling the differences in counted outcomes using bivariate copula models with application to mismeasured counts. Econometrics Journal 7(2), 566–584.

Casagrande, S.S., Wang, Y., Anderson, C., Gary, T.L., 2007. Have Americans increased the fruit and vegetable intake? American Journal of Preventive Medicine 32(4), 257–263.

Cox, T.L., Wohlgenant, M.K., 1986. Prices and quality effects in cross-sectional demand analysis. American Journal of Agricultural Economics 68(4), 908–919.

FAOSTAT, Food and Agriculture Organization Statistics Division, 2009. Accessed June 3, 2009, available at http://faostat.fao.org.

Feng, X., Chern, W.S., 2000. Demand for healthy food in the United States. Paper presented at AAEA annual meeting, Tampa, FL, 30 July–2 August.

Gouriéroux, C., Monfort, A., Trognon, A., 1984. Pseudo maximum likelihood methods: applications to Poisson models. Econometrica 52(3), 701–720.

Gustavsen, G.W., Rickertsen, K., 2002. Public policies and the demand for vegetables. Paper presented at the Xth EAAE Congress, "Exploring Diversity in the European Agri-Food System", Zaragoza, Spain, 28–31 August.

Gustavsen, G.W., Rickertsen, K., 2006. A censored quantile regression analysis of vegetable demand: the effects of changes in prices and total expenditure. Canadian Journal of Agricultural Economics 54(4), 631–645.
Huang, K., 1993. A complete system of U.S. demand for food. Technical Bulletin No. 1821, U.S. Department of Agriculture, Economic Research Service, Washington, DC, September, 70 pp.
Huang, K., Lin, B., 2000. Estimation of food demand and nutrient elasticities from household survey data. Technical Bulletin No. 1887. U.S. Department of Agriculture, Economic Research Service, Washington, DC, August, 30 pp.
Kocherlakota, S., Kocherlakota, K., 1993. Bivariate Discrete Distributions. Marcel Dekker, New York.
Lee, J., 1985. The demand for varied diet with econometric models for count data. American Journal of Agricultural Economics 69(3), 687–692.
Lee, L-F., 1983. Generalized econometric models with selectivity. Econometrica 51, 507−513.
Marshall, A.W., Olkin, I., 1990. Multivariate distributions generated from mixtures of convolution and product families, in H.W. Block, A.R. Sampson, and T.H. Savits, eds., Topics in Statistical Dependence, IMS Lecture Notes-Monograph Series, vol. 16, pp. 371–393.
Ministry of Health Malaysia, 2006. Malaysia NCD Surveillance 2006: NCD Risk Factors in Malaysia. Ministry of Health, Disease Control Division, Malaysia.
Nelsen, R.B., 2006. An introduction to Copulas. Springer, New York.
Stewart, H., Harris, J.M., Guthrie, J., 2004. What determines the variety of a household's vegetable purchases? Agriculture Information Bulletin Number 792–3. U.S. Department of Agriculture, Economic Research Service, Washington, DC, November, 4 pp.
U.S. Department of Health and Human Services (USDHHS) and U.S. Department of Agriculture (USDA), 2005. Dietary Guidelines for Americans, 2005, 6th Edition, U.S. Government Printing Office, Washington, DC, January.
Vuong, Q.H., 1989. Likelihood ratio tests for model selection and nonnested hypotheses. Econometrica 57(2), 307–333.
World Health Organization (WHO), 2003. World Health Report 2002: Reducing Risks, Promoting Healthy Life. World Health Organization, Geneva. Accessed June 2, 2009. available at http://www.who.int/whr/2002/en/whr02_en.pdf.
You, Z., Epperson, J.E., Huang, C.L., 1998. Consumer demand for fresh fruits and vegetables in the United States. Research Bulletin Number 43, The Georgia Agricultural Experiment Stations, The University of Georgia.

Table 1 Definitions and sample means of explanatory variables (*n* = 2,447)

Variable	Definition	Mean
Continuous variable		
Work hours	Length of typical work day (hours)	7.40
		(3.17)
Binary variables (1 = yes; 0 = no)		
Primary	Primary as highest level of education (reference)	0.42
Junior high	Junior high as highest level of education	0.22
Senior high	Senior high as highest level of education	0.26
Tertiary	Tertiary as highest level of education	0.10
Age ≤ 30	Age is 30 years or below (reference)	0.13
Age 31–40	Age is between 31 and 40 years old	0.27
Age 41–58	Age is between 41 and 58 years old	0.51
Age ≥ 59	Age is 59 years or above	0.09
Malay	Ethnicity is Malay	0.55
Chinese	Ethnicity is Chinese	0.18
Indian	Ethnicity is Indian	0.09
Others	Ethnicity is one of others (reference)	0.18
Poverty-income	Monthly household income is RM0 – 399	0.11
Low-income	Monthly household income is RM400 – 999	0.36
Middle-low income	Monthly household income is RM1000 – 2999	0.38
Middle-high income	Monthly household income is RM3000 – 5999	0.06
High-income	Monthly household income RM6000 or above (reference)	0.03
Male	Gender is male	0.41
Single	Marital status is single, divorced or widowed	0.13
Rural	Reside in rural area	0.50
Smoker	Currently smoking cigarettes	0.21
Hyperchol	Diagnosed with hypercholesterolemia	0.56
High BP	Diagnosed with high-blood-pressure	0.32
Diabetes	Diagnosed with diabetes	0.13
Region1	Penang, Selangor, Federal Territory	0.19
Region2	Perlis, Kedah, Perak, Melaka, Negeri Sembilan, Johor, Pahang, Kelantan, Terengganu (reference)	0.41
Region3	Sabah, Sarawak	0.40

Note: Compiled from Ministry of Health Malaysia (2006). Standard deviations are in parentheses. As of 2 June 2009, exchange rate was approximately US$1.00 = RM3.51. The five income categories correspond to poverty (US$0–113.90), low (US$114–284.90), middle-low (US$285–853.90), middle-high (US$854–1708.90) and high (≥ US$1709).

Table 2 Maximum-likelihood estimates of bivariate count system for fruit and vegetable consumption: days per week

	Fruits		Vegetables	
Variable	Estimate	S.E.	Estimate	S.E.
Constant	1.045***	0.085	1.874***	0.083
Work hours	0.005	0.004	0.002	0.004
Junior-high	0.077**	0.035	0.036	0.031
Senior-high	0.191***	0.036	0.051	0.035
Tertiary	0.169***	0.050	0.056	0.055
Age 31–40	0.053	0.046	–0.027	0.041
Age 41–58	0.191***	0.046	0.018	0.040
Age ≥ 59	0.243***	0.063	0.014	0.056
Malay	0.140***	0.050	–0.048	0.051
Chinese	0.189***	0.053	–0.019	0.058
Indian	0.180***	0.064	0.018	0.063
Poverty-income	–0.295***	0.061	–0.145**	0.058
Low-income	–0.203***	0.048	–0.129***	0.050
Middle-low income	–0.080*	0.046	0.011	0.051
Middle-high income	–0.070	0.067	0.032	0.080
Male	–0.041	0.031	–0.046	0.031
Single	0.012	0.037	–0.034	0.035
Smoker	–0.087**	0.037	–0.026	0.034
Hyperchol.	0.063**	0.027	0.014	0.025
High BP	0.011	0.029	–0.018	0.027
Diabetes	–0.026	0.038	–0.039	0.034
Region 1	–0.096***	0.034	–0.222***	0.028
Region 3	0.192***	0.044	0.081	0.050
α	0.345***	0.062		
ρ	0.313***	0.023		
Log likelihood	–10410.439			

Note: *p < = 10%; **p < = 5%; ***p < = 1%.

Table 3 Marginal effects of explanatory variables on the mean days of servings per week: bivariate count model for days per week

Variable	Fruits	Vegetables
Continuous explanatory variables		
Work hours	0.018	0.012
	(0.015)	(0.023)
Binary explanatory variables		
Junior-high	0.273**	0.214
	(0.126)	(0.185)
Senior-high	0.724	0.300
	(0.140)	(0.206)
Tertiary	0.630***	0.331
	(0.194)	(0.332)
Age 31–40	0.179	–0.160
	(0.152)	(0.238)
Age 41–58	0.688***	0.107
	(0.155)	(0.239)
Age ≥ 59	0.898***	0.084
	(0.238)	(0.331)
Malay	0.495***	–0.287
	(0.171)	(0.310)
Chinese	0.684***	–0.113
	(0.186)	(0.349)
Indian	0.648**	0.112
	(0.233)	(0.387)
Poverty-income	–1.096***	–0.847**
	(0.229)	(0.343)
Low-income	–0.787***	–0.758**
	(0.198)	(0.307)
Middle-low income	–0.329*	0.069
	(0.196)	(0.320)
Middle-high income	–0.289	0.202
	(0.274)	(0.510)
Male	–0.153	–0.268
	(0.114)	(0.181)
Single	0.046	–0.197
	(0.140)	(0.200)
Smoker	–0.318**	–0.150
	(0.133)	(0.201)
Hyperchol.	0.234**	0.080
	(0.099)	(0.145)
High BP	0.042	–0.106
	(0.108)	(0.160)
Diabetes	–0.096	–0.229
	(0.141)	(0.194)
Region 1	–0.331***	–1.208***
	(0.115)	(0.148)
Region 3	0.768***	0.512
	(0.185)	(0.322)

Note: Standard errors in parentheses. **p* < = 10%; ***p* < = 5%; ****p* < = 1%.