Factors Affecting Malnutrition in Developing Countries: A Linear Mixed Effect Model Approach

Sohair F. Higazi
Dept. Statistics and Insurance
Faculty of Commerce, Tanta University
Egypt
sohair.higazi@gmail.com

Dina Hassan Abdel-Hady
Dept. Statistics and Insurance
Faculty of Commerce, Tanta University
Egypt
dina1002007@yahoo.com

Samir A. M. Al-Oulfi
Dept. of Statistics and Insurance
Faculty of Commerce, Mansoura University
Egypt
samiraloulfi@yahoo.com

Abstract

The main objective of this study is to pinpoint the main factors that affect the percentage who suffers of malnutrition in developing countries. Three locations are randomly chosen: Asia, Africa, and Middle east and North Africa (MENA); A total of 96 countries were chosen randomly from 137 developing countries of the three locations; and were cross classified by "Location" and "Human Development Index (HDI) as high, middle, and low (UNDP, 2005)[i]. Data for the study was compiled from FAO (2005)[ii]. The analysis started with seven explanatory variables and the dependent variable; however, stepwise regression reveals that the average Protein intake and Infant mortality rate were the only two significant variables. "Location and "HDI" are dummy coded and OLS regression is performed using the two significant variables, but the only significant variable was the "average protein intake". OLS multiple regression Model is re-applied to the data using dummy variables technique with interaction with the "average Protein intake", nine regression equations were reached.

The Linear Mixed effect Models are also applied, using "location" as the random factor and "HDI" as the fixed factor. Five models were applied: (1) a null model (baseline model)where no predictors are introduced to the model; (2) the fixed model: where predictors used are the covariate and the HDI; (3) the random model: where predictors used are the covariate and Location; (4) the mixed model: where predictors used are the covariate and the HDI I (fixed) and the location (random); and (5) the random coefficient model: where predictors used are the covariate, the HDI Index and the location but produces different prediction equations that differ in slopes and intercepts. Models are compared based on information criterions. The random coefficient model produces the least criterion values and thus fits better than all previous ones. A comparison between the Random Coefficient model results and GLM model is made, and conclusions are reached.

Keywords: Linear Mixed Models, Hierarchical Models, Null Model, Random Coefficient Model, GLM, OLS regression, Malnutrition Developing Countries.

1. Introduction

Malnutrition is a major problem in developing countries, especially with the decrease of food supply and accordingly the increase of food prices all over the globe. Developing countries suffer from many social and economic problems such as: high birth rate, high fertility rates, high unemployment rates, low average productivity and low GDP per person. Almost all developing countries[iii] suffer from malnutrition. Malnutrition essentially means "bad nourishment", it could be "over nutrition" when the diet contained too many calories, or could be "under nutrition" if the diet does not provide adequate calories and protein for the growth and body maintenance, called" Protein –Energy Malnutrition". In this study we define malnutrition as "under nutrition". Malnutrition" plays a major role in half of all under five deaths each year in developing countries (WHO 2000)[iv]. It is mainly caused by food insecurity, lack of clean water, and insufficient household income (Setboonsarng2005)[v]; Chronic food insufficiency affect about 792 million persons in the world (FAO 2000), that includes 20% of the population in developing countries, it affects all age groups, especially common among the poor; about 70% of children with "Protein-Energy Malnutrition" are in Asia, 26% live in Africa; 4% in Latin America and the Caribbean (WHO 2000)[vi]. By 2015, 682 million persons are expected to suffer malnutrition. The number of persons (in million) who suffers malnutrition in developing countries in selected years is given in the following table[vii]:

No. of persons suffer from malnutrition in millions by location			
Location	1990-1992	1995-1997	1999-2001
Asia and the Pacific	567	496	505
Latin America and the Caribbean	59	55	53
Middle east-North Africa	25	35	41
Africa Sub-Saharan Dessert	166	193	198
Developing countries	**817**	**780**	**798**

In developing countries, protein-energy consumption has increased from 57 kg/person in 1960-1970 to 75 kg/ person in 1990 (FAO[viii] 2003) while it is one-third of that in industrial countries who suffer from "Over-nutrition". Calorie consumption from dairy products and protein are increasing in Latin America and the Caribbean, average calorie/ person consumption is 550 Calories (SIWI et al. 2005, FAO 2004) while in China, Kcal increased from 100 Kcal in 1961 to 400 Kcal in 2001 In Asia, Kcal increased from 100 Kcal in 1961 to 370 Kcal in 2001. In Africa, Kcal is constant (FAO 2003). Globally, calorie consumption from animal products has increased from 100 Kcal in 1961 to 200 Kcal in 2001 (FAO 2004, SIWI et al. 2005)[ix].

Several researchers have studied malnutrition in developing countries; in 2004 the Bangladesh Bureau of Statistics (United nation World Food Program 2004) has conducted a comprehensive study about poverty and malnutrition using 63 predictors. The research depended on stratification using GIS data and used step wise regression to reduce the number of predictors. Several regression lines were produced, and log (expenditure) was the used as "the" covariate that produces several regression lines with different intercepts, slopes and R^2 value for each stratum. Child malnutrition was studied in Indonesia (ADB 2001, Setboonsang 2005, Atmarita et al 2000a, b, World Bank 1994).

Malnutrition was used as a poverty indicator; the study was descriptive in nature. Analysis of some malnutrition cases was performed in Zimbabwe (James 2002a, b). Webb and Lapping (2002) conducted a study for food and nutrition insecurity in China, Egypt, Ghana, Indonesia, Myanmar, and South Africa; the study aimed to study the relationship between nutrition, food, income and other dimensions of household livelihood. Webb and Lapping used a multistage stratified random sampling process; focus groups were formed when no qualitative data were available. Only in Indonesia, Myanmar and Ghana econometric regression analysis was performed to determine relationships among variables, the statistical methods were largely descriptive supplemented by analysis of variance (ANOVA).

The main objective of the present study is to reach factors affecting malnutrition in developing countries. Variables used for the analysis include: average GDP, GAP, illiteracy rate, average protein intake, ratio of dairy production to consumption, percent who suffer from malnutrition and infant mortality rates. The response variable used is the percentage of persons suffers from malnutrition, as measured by FAO[x] (2005). Three *locations* are randomly chosen, they are: Asia, Africa, and Middle East and North Africa (MENA); 97 countries were chosen randomly from countries of the three locations, and were classified according the area and also according to the *Human Development Index* (*HDI)*: high, middle, and low development countries. The "Location" and "*HDI*" factors were chosen to reflect "general" economic, social, and environment conditions that are common within countries in each combination of "location" and "HDI".

To reach the above objective, many statistical techniques may be used. The general linear model *(GLM*), which takes the form: $Y = \mathrm{XB} + \mathrm{E}$, the explanatory variables (X's) are assumed to be fixed values, it could be quantitative in nature and could be also categorical that have been dichotomized by using dummy variable technique. The GLM includes analysis such as correlation, t-tests, analysis of variance, regression, etc**.** The present study uses Linear Mixed Models (LMM) which is a generalization of GLM, and used for the analysis of a continuous dependent and "fixed" and "random" effects in a hierarchical design (called also "Multilevel Models MLM or Hierarchical Mixed Models HMM").In hierarchical models, observations (countries in our study) at one level will share same history, and this clustering increase type I error, but LMM does take this inter-class dependence into consideration. The random effects do not affect population means but they do affect the covariance structure of the data and, indeed, adjusting for this is a central point of LMM models and why they are used instead of GLM, which assumes independence.[xi]

In section 2, we introduce the mixed models and the estimation algorithm, and in Section 3, we give the methodology followed, and in Section 4, data analysis results are presented; recommendations are given in Section 5.

2. The General Form of the Mixed Effects Model

The application of Mixed Linear Models (MLM) gives different conclusions when compared to conventional regression analysis. The dependent variable in multilevel modeling *is a* normally distributed quantitative variable (percent of persons suffers from malnutrition), it is linearly related to the fixed (HDI), to the random factors (Location) and to the covariates. These models are used to reveal the correlation between correlated levels of the fixed and the random levels. The fixed factor(s) are looked at as "Blocking factors", the random factor(s) is a random sample of all levels of the random factor. Mixed models are good when researcher is interested invariance components (Kreft1995); Random coefficient models, as MLM provide researchers with separate estimates for separate levels.

The main difference between MLM and OLS (Moerbeek, van Breukelen, and Berger 2003[xii]) is that OLS gives incorrect standard error for treatment effects, and thus, "multilevel" regression should be used. In addition, random effects which could result from clustering observations, (for example, grouping countries by HDI), this clustering increase type I error and leads to correlated error terms and biased estimates which violates the OLS assumptions[xiii]; these violations are handled by MLM,since it does take this inter-class dependence into consideration. However, the formed groups (blocks) are assumed to be independent and to have the same covariance structures, this is an advantage of LMM models and why they are used instead of GLM, which assumes independence. The error variance is assumed to be constant across the formed groups, unlike OLS regression.

The coefficients in Linear Mixed Models (LMM) are looked at as "random" effects drawn from a normal distribution of possible coefficients, whereas OLS regression treats the coefficients' parameters as if they were "fixed constants". In addition, LMM can handle a random sampling variable like "Location" even when there are too many locations, which require to be dichotomized into dummy variables in OLS regression.

The linear mixed population model takes the form (Nyoman Latra et al., 2010):

$$y = X\beta + Z\gamma + \varepsilon$$

Where:

y : is a $n \times 1$ vector of observations,

X : is a $n \times (p+1)$ matrix of observation from fixed known covariates,

Z : is a $n \times m$, containing m variables of random effects.

β: is a $(p + 1) \times 1$ vector of unknown regression coefficients which are usually known as the fixed effects,

γ : is a $m \times 1$ vector of random effects, and

ε: is a $n \times 1$ vector of random errors.

Both vectors **γ** and **ε** are unobservable. The basic assumption of the above equation for **γ** and **ε** are:

1. $\boldsymbol{\gamma} \sim N(\mathbf{0},\mathbf{D})$
2. $\boldsymbol{\varepsilon} \sim N(\mathbf{0}, \boldsymbol{\sigma}^2)$

Where **D** and $\boldsymbol{\sigma}^2$ are the variances- covariance matrices[xiv].

The main assumption in the mixed model is that γ and ε are both normally distributed with:

$$E\begin{bmatrix}\gamma\\ \varepsilon\end{bmatrix}=\begin{bmatrix}0\\ 0\end{bmatrix} \quad \text{And} \quad Var\begin{bmatrix}\gamma\\ \varepsilon\end{bmatrix}=\begin{bmatrix}D & 0\\ 0 & \sigma^2\end{bmatrix}$$

The matrix D is a 2x2 covariance matrix of the random effects**:**

$$D=\begin{bmatrix}\sigma^2{}_{\gamma 1} & \sigma_{\gamma 1,\gamma 2}\\ \sigma_{\gamma 1,\gamma 2} & \sigma^2{}_{\gamma 2}\end{bmatrix}$$

While the fixed effect matrix is diagonal:

$$\sigma^2=\begin{bmatrix}\sigma_\varepsilon{}^2 & 0 & 0 & 0 & \dots & 0\\ 0 & \sigma_\varepsilon{}^2 & 0 & 0 & & 0\\ 0 & 0 & \sigma_\varepsilon{}^2 & 0 & . & 0\\ 0 & . & . & . & & \\ 0 & 0 & 0 & 0 & \dots & \sigma_\varepsilon{}^2\end{bmatrix}=\sigma_\varepsilon{}^2 I$$

The resulting estimated variances of y given X and Z is:

$$Var(y|X,Z)=\sigma^2+ZDZ^T$$

and the abov*e* equation is written as:

$y=XB+\varepsilon^*$ **and y** ~ $N(XB,\sigma^2+ZDZ^T)$ (Nyoman Latra *et al*., 2010)**.** Parameters to estimate are: β and any parameter in D. If D is diagonal, then there are m "variance" components. The covariance matrix of y is then a "block" diagonal with diagonal matrices $\boldsymbol{\sigma}^2+\mathbf{ZDZ}^T$ and off diagonal matrices of zero[xv].

2.1 Estimation of β, $\Sigma_{\gamma,\,\Sigma}$

The variance components technique is used for the estimation of the variance-Covariance matrix (Garson 2008). The "Variance component" does not assume any correlation between the random effect factor levels; it is accompanied by an identity matrix with each random effect factor, called "Random Effect Covariance Structure" (Muthén, B.O., 1994) which takes the form $\boldsymbol{\sigma}^2\mathbf{I}$ (called the **G** matrix in SPSS). The covariance structure matrix is a table in which both rows and columns are the values of the grouping "random" variable which is Location (Asia, Africa, and MENA). Cell entries represent the covariance of the residuals when predicting the dependent variable, using all countries.

This Random Effect Covariance structure is the initial starting point for the estimation process. LMM uses an iterative algorithm to estimate coefficients. Two iterative methods may be applied: the **Re**stricted **M**aximum **L**ikelihood Estimation (**REML)** and **M**aximum **L**ikelihood Estimation **(ML**). REML handles high correlations more effectively, and is less sensitive to outliers than ML, but cannot be used for model comparison of fixed effects; its estimates are the same as ML estimates for large samples[1]

Maximum likelihood(ML) estimation finds the parameter estimates (regression coefficients) which maximize the probability that the estimate of the dependent variable equal to the observed values, but ML estimates ignore the degrees of freedom used up by fixed effects in mixed models, leading to underestimation of variance components. REML is a "residual" maximum likelihood, since the likelihood function includes only variance components, while ML includes regression coefficient. The REML and ML estimation methods give asymptotically efficient estimates for unbalanced as well as balanced designs (Hoxand Maas 2001).

The REML does not base estimates on a maximum likelihood fit of all information, but on a likelihood function calculated from a transformed set of data (Dodge2003). In case of variance components, estimation is based on a set of contrasts and the likelihood function is calculated from a probability distribution based on the formed contrasts, REML produces unbiased estimators of the variance-covariance parameters (Searle, Casell, and McCulloch 2008).

2.2 Goodness of Fit Statistics

Five goodness of fit measures may be used for the comparison of linear mixed models(Ritz, 2004)[xvi]: -2 REML , Akaike Information Criterion (AIC), Hurvich and Tsai criterion which is AIC Corrected, for finite sample corrected (AICC); Bozdogan's criterion which is a consistent AIC (CAIC) and Schwarz's Bayesian Criterion (BIC). These last four criterions are:

1. Akaike Information Criterion:
   $$AIC = -2\log LL + 2n$$
2. Akaike Information Criterion corrected:
   $$AICC = -2LL + 2\frac{pn}{n-p-1}\;)$$
3. BozdoganAkaike Information Criterion consistent:
   $$CAIC = -2LL + p\left(\log(n) + 1\right)$$
4. Bayes Information Criterion**:**
   $$BIC = -2\log(LL) + nLog(N)$$

Where n is the number of parameters in the model, and N is the total number of observations.

Simulation studies suggest BIC is preferred generally, but AICC is better for sample sizes < 20 and when there is only one group(Ehlers2004)[xvii].

The F test is used to test the fitness of the model to the data and to test the significance of the R^2 value (Laird, N. M. and J. H. Ware1982).When fitting more than one model to the data, the four criterions are compared; the model with the least value is considered the model that gives the best fit. The likelihood Ratio test is also used for the comparison of the models. Where the difference between -2LL in the two compared models is computed; i.e.

$$Diff = -2LL_1 - (-2LL_2)$$

Where: LL_1 is log likelihood from model 1 and LL2 is the log likelihood from model 2; the difference is compared to a chi-square value with parameters equal to the difference in parameters in the two compared models.

3. Methodology

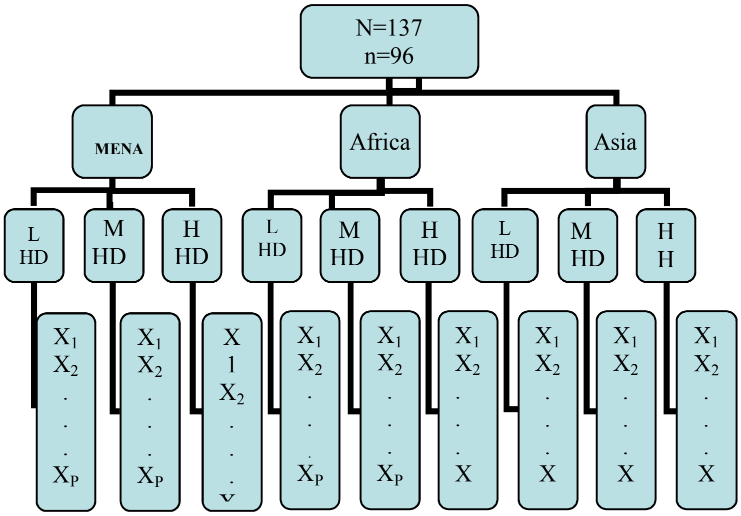
To run the Linear Mixed Model, we distinguish between the following models[xviii]:

1. The null model: predicts the dependent variable from the intercept and error term only, i.e.; no predictors. Thus we are testing $H_0 : \hat{y} = 0$ against $H_1 : \hat{y} = \bar{y}$ this model is considered a base line for comparisons between models, information criterions produced from this model is the baseline criterion when introducing fixed and random factors.
2. The fixed effect model: predicts the dependent variable from the covariate, fixed factors and error term. Thus we are testing $H_0 : \hat{y} = \bar{y}$ against $H_1 : \hat{y} = XB$, goodness of fit criterions are compared to those of the null model.
3. The random effect model: predicts the dependent variable from the intercept, random factors and optional covariates, and error term. Thus we are testing: $H_0 : \hat{y} = \bar{y}$ against $H_1 : \hat{y} = Z\hat{\gamma}$, goodness of fit criterions are compared to those of the null model.
4. The mixed model or the hierarchical model: has the fixed factors, the random factors and optional covariates. Thus we are testing: $H_0 : \hat{y} = X\hat{B}$ against $H_1 : \hat{y} = X\hat{B} + Z\hat{\gamma}$, goodness of fit criterions are compared to those of the null model.
5. The random coefficient (RC) model: is a type of HLM when it is assumed that each group at a higher level has different slopes and different intercepts for the prediction of the dependent variable, which is similar to dummy variable technique with interaction used in OLS, however some effects are considered fixed and others are random**.**

Hox (1995) suggests proceeding testing the above models in sequential steps, at any test to see if the model has a significantly better fit than the baseline model and of the preceding model. If it does, proceed to investigate the next model. At each step identify which base-level regression slopes have significant variance across upper level groups. Drop variables which do not improve model fit. Add cross-level interactions between explanatory variables and base level independent variables that had slope variance. Drop interactions which do not improve model fit.

4. Data Analysis Results

The main objective of this paper is to pinpoint the main factors that cause malnutrition (MALN %) in developing countries using linear mixed models. Three locations are chosen in random, they are: Asia, Africa, and Middle east and North Africa (MENA); Available data were obtained from FAO (2005) for 137 countries 96 countries were chosen randomly from countries of the three locations, and then classified according the "Location" and also according to the Human Development Index (UNDP2005) as high, middle, and low development countries. The design is as follows:



Seven explanatory variables were used in the analyses that include: illiteracy rates, GDP, Agriculture Domestic Product (ADP), Poverty ratios, infant mortality rates, average protein intake, dairy production to consumption. Descriptive measures for the variables used in the analysis are exhibited in Table (1).

It is evident from Table(1) that the highest percentage of malnutrition is in the low HDI of the Asian countries (61%), the next highest is in High HDI countries of the Middle East and North African Countries (30%).Also, the highest protein consumption is in the high HDI of the MENA countries (105 gm/day), and the lowest protein daily consumption is in the low HDI country of Asia (Afghanistan).

Pearson's correlation coefficients were computed, correlation coefficients between the dependent variable (MALN %) and all other predictors are significant, but insignificant between MAlnut% and Poverty ratios P=.074). Thus poverty ratio was excluded, and a stepwise multiple regression analysis (n=96, k=6) is applied, the only significant predictors

Table (1): Descriptive measures for variables used in the analysisby Location and HD Index

Location	n	HDI	Malnu %	Pov. rate	Average Protein(gm)	Infant Mort. Per 000	Illit. rate	GDP $	ADP $	Dairy Prod/ Cons. %
Asia	**5**	**H**	20	32	87	14	4	9796	9719	98
	24	**M**	18	27	66	57	20	967	449	77
	1	**L**	61	32	51	140	90	210	90	43
Africa	**2**	**H**	11	32	78	21	26	4973	691	10
	12	**M**	19	45	61	100	22	1780	472	43
	31	**L**	35	47	53	180	43	341	152	75
MENA	**5**	**H**	30	20	105	11	9	17427	32580	47
	14	**M**	8	16	75	34	18	3170	2428	90
	2	**L**	22	32	66	96	26	934	2139	70

were: Infant mortality rates and the average daily protein intake. The OLS equation *for* predicting MALN% is:

$$MALN\ \% = 0.498 - .005\,(protein\ ) + .001\,(INF\ )$$

$$R^2 = 53.2\% \qquad n = 96$$

Thus; percent of malnutrition is related inversely with the amount of daily protein intake, and is related directly with infant mortality rate, for example, Egypt's infant mortality rate is .039 and the average protein daily intake is 93 grams; thus, the percent of malnutrition is expected to be 12.6%, and a country asTajikistan, with infant mortality rate of 118 per thousand and average protein intake is 48 gm/day, the average malnutrition is expected to be 37.6%. It is evident that the above estimates have high error of estimate.

4.1 Results of OLS with Dummy Variables

Applying OLS regression with dummy variables for "Location" reveals the significance of the covariate (average protein consumption) only. Applying the model with interaction term with "average protein consumption" gives an adj-R^2 value=61.3%, equation for each combination of HDI and location are given in Table (2).

One country was chosen from each of eight combinations of location and HDI (the only country in Asia and Low HDI was excluded). Using equations in Table (2), the percentages of malnutrition are estimated for each selected country and given in Table (2-a).

Table (2): Prediction Equations for Malnutrition by Location and HD Index

Location	HDI	$\hat{y}$:Average Predicted MALN%
Asia	H	$\hat{y} = .747 - .007 \times Protein$
	M	$\hat{y} = .421 - .003 \times Protein$
	L	$\hat{y} = .747 - .001 \times Protein$
Africa	H	$\hat{y} = .837 - .009 \times Potein$
	M	$\hat{y} = .511 - .005 \times Protein$
	L	$\hat{y} = .947 - .011 \times Protein$
MENA	H	$\hat{y} = .411 - .004 \times Protein$
	M	$\hat{y} = .411 - .004 \times Protein$
	L	$\hat{y} = .411 - .004 \times Protein$

Table (2-a): Estimated percent of Malnutrition using Dummy Variables technique for some selected Countries

Country	Location/HDI	%Malnut.	Protein intake	Estimated %Malnut.
Korea	Asia/H	35	63	30.6
Tajikistan	Asia/M	61	48	27.7
Cecile	Africa/H	.19	84	8.1
Congo	Africa/M	.34	43	29.6
Chad	Africa/L	.33	66	22.1
Kuwait	MENA/H	.50	84	7.5
Egypt	MENA/M	.30	93	3.9
Yemen	MENA/L	.37	57	18.3

It is evident that the model above underestimates the percent of malnutrition in all selected countries.

4.2 Results of Mixed Model

The mixed model is applied using average protein consumption as a covariate, HDI as the fixed factor, coded as "1" for "High" HDI, and "2" for "Middle" HDI, and "0" for "Low" HDI. The "Location" is the random factor, coded as "1" for" Asia", "2" for "Africa", and "0" for "MENA". Following the steps suggested by Hox (1995), the null (baseline), the fixed effect and the random effect models were performed. The null model estimates only the intercept, it shows that, on the average 22% of all developed countries suffer from malnutrition, the percentage ranges from 18.5% at 25.43% in 95% of all developing countries. The fixed effect model with interaction term gives an estimate for the covariate, and the fixed factor levels and for the interaction between the covariate and each level of the fixed factor, as shown in Table (3).

Table(3): Estimates of Fixed Effects

						95% CI	
Parameter	Estimate	Std. Error	df	T	Sig	Lower	Upper
Intercept	.9456	.0876	90	10.79	.000	.7715	1.1198
HDI=1	-.3662	.1927	90	-1.90	.061	-.7490	.0165
HDI=2	-.4689	.1151	90	-4.07	.000	-.6976	-.2402
Protein	-.0111	.0016	90	-6.93	.001	-.0143	-.0079
HDI=1 * Protein	.0061	.0024	90	2.52	.014	.0013	.0110
HDI=2 * Protein	.0065	.0019	90	3.41	.001	.0027	.0103

The prediction equation is: $\hat{y} = .5794 - .005(Protein)$ *for High Index countries and* $\hat{y} = .477 - .005(\mathrm{Pr}otein)$ for *Middle" Index countries and* $\hat{y} = .946 - .0111(\mathrm{Pr}otein)$ *for Low" Index countries. Using the prediction, equation, the expected percent for malnutrition for Korea is 26.44%, 15.94% for Cecile and Kuwait, 23.7% for Tajikistan, 26.2% for Congo, 21.34% for Chad, and 31.33% for Yemen.*

The random effect model produces only an intercept; no estimates were given for the covariate or the random effect (Location). The "G" matrix is obtained by "Variance Component" method, it equals 0.0156. The information criterions produced from the three models are given in Table (4).

Table (4): Information Criterions for the Null, Fixed and Random Models

Criterion	Null Model	Fixed Model	Random Model
-2 RELM	-62.417	-114.092	-93.161
AIC	-60.417	-112.092	-89.161
AICC	-60.374	-112.047	-89.031
CAIC	-56.863	-108.592	-82.053
BIC	-57.863	-109.592	-84.053

Comparing criterions in Table (4) reveals that both the fixed and the random models fit the data better than the null model; but the fixed model shows better fit than the random model. The likelihood ratio test for testing the null model against the fixed model is:-62.417-(-114.092)=51.675, this statistic has a chi-square distribution with five degrees of freedom (the difference in parameters in the two models), this gives a significant chi-square value (α=5%).

4.3 The basic Linear Mixed Model

This model includes: "fixed" effect factor (the covariate: Protein and HDI: High, Middle and Low); and the "random" effect factor (Location: Asia=1, Africa=2, MENA=3) in the analysis. Subjects are "countries" within each combination of Location and HDI. Using SPSS (Mixed Model), the following estimates in Table (5) are obtained:

Table(5): Estimates of Fixed Effects Dependent variable: % malnutrition

						95% CI	
Parameter	Estimate	Std. Error	df	t	Sig	Lower	Upper
Intercept	.6858	.0490	10.397	13.984	.000	.5771	.7945
HDI=1	.0087	.0498	10.165	.175	.864	-.1020	.1195
HDI=2	-.0943	.0294	1.446	-3.210	.127	-.2807	.0920
Protein	-.0062	.0008	54.840	-7.470	.000	-.0079	-.0046

The output does not contain any estimates for the random effect factor, however estimate for the intercept and for the covariate are smaller than those given in Table (3). The prediction equation for High HDI is: $\hat{y} = .6945 - .006(Protein)$; for " Middle" HDI is: $\hat{y} = .5915 - .006(\mathrm{Pr}otein)$; and for " Low" HDI countries: $\hat{y} = .6858 - .006(\mathrm{Pr}otein)$. This model gives an estimate of 31.65% for Korea, 13.65% for Egypt, 33.35% for Congo, and 33.24% for Yemen. Thus, the basic model gives a smaller error of estimate than the dummy variable technique model.

The "G" matrix for the random effects is obtained by "variance Component" variance structure (the G matrix) shown in Table (6).

Table (6): Random Effect Covariance Structure (G)

	Location=1 | Location *HDI	Location=2 | Location *HDI	Location=3 | Location *HDI
Location=1 | Location *HDI	5.30566E-005	0	0
Location=2 | Location *HDI	0	5.30566E-005	0
Location=3 | Location *HDI	0	0	5.30566E-005

Variance components, Dep. Variable: % Malnut.

The variance component structure does not assume any correlation between the random effect factor levels (Asia, Africa and MENA). Thus the variance for each location, controlling for HDI is constant = 0.00005**I**, where **I** is an identity matrix.

4.4 The Random Coefficient Model

To overcome the problem of not reaching estimates for the random effect factor, we used the Random Coefficient model (Table 7). The random factor "Location" is not significant, however, it is a sort of "dummy" variable, that is needed only for the adjusting of the prediction equation for each combination of" Location" and HDI. All equations will have same slope coefficient for the covariate "Protein" but different intercepts depending on the location and the HDI.

For example, the prediction equation for the Asian countries with high HDI is: $\hat{y} = .7209 - .0061(\text{Protein})$which yields an estimate of 33.66% for Korea (as an example), and for the Asian countries with Middle HDI, the equation is: $\hat{y} = .6178 - .0061(\text{Protein})$, this yields an estimate of 10.54% for Cecile (as an example), and for MENA countries with high HDI, the equation is: $\hat{y} = .6708 - .0061(\text{Protein})$, this yields an estimate of 15.84% for Kuwait (as an example) and for MENA countries with low HDI, the equation is: $\hat{y} = .6974 - .0061(\text{Protein})$, this equation yields an estimate of 34.97% for Yemen. These estimates for the above selected countries have less error of estimates than the previous models presented.

Table (7): Estimates of Fixed* Effects

						95% CI	
Parameter	Estimate	Std. Error	Df	t	Sig	Lower	Upper
Intercept	.6974	.0751	90	9.289	.000	.5482	.8465
Location=1	.0501	.0337	90	1.487	.141	-.0168	.1171
Location=2	-.0193	.0384	90	-.503	.616	-.0956	.0569
HDI=1	-.0266	.0511	90	-.521	.604	-.1281	.0749
HDI=2	-.1297	.0329	90	-3.938	.000	-.1952	-.0643
Protein	-.0061	.0009	90	-6.674	.000	-.0080	-.0043

*Dep. Var. % Malnutrition

The information criterion for the mixed model *and the Random Coefficient (RC) models are given in Table (8).*

Table (8): Information Criterion for the Mixed and the RandomCoefficient Models

Criterion	**Null Model**	**Mixed Model**	**RC Model**
-2 RELM	-62.417	-118.649	-123.670
AIC	-60.417	-114.649	-119.670
AICC	-60.374	-114.511	-119.536
CAIC	-56.863	-107.649	-112.627
BIC	-57.863	-109.649	-114.627

Comparing information criterions in Table (8) with those in Table (4) shows that the mixed model has a better fit to the data and the RC model gives the best fit. The likelihood ratio test for testing the null model against the mixed model is:-62.417-(-118.649)=56,232, this statistic has a chi-square distribution with threedegrees of freedom, this gives a significant chi-square value (α=5%). The likelihood ratio test for testing the null model against the random coefficient model (RC) is:-62.417-(-123.670)=61.253, this statistic has a chi-square distribution with fivedegrees of freedom, this gives a significant chi-square value (α=5%). The likelihood ratio test for comparing the mixed model to the RC model is 5.021 which have a chi-square distribution with 2 degrees of freedom ($\chi 2_{.05} = 5.99$); thus the mixed model does not differ significantly from the RC model, and both models differ significantly than the null model.

4.5 The Random Coefficient Model With Interaction

To test if there is interaction between the covariate (Average Protein intake) and each level of the fixed and random factors, the mixed model is used with interaction, the following Table of estimates is obtained (Table (9)).

Table(9): Estimates of Fixed* Effects

						95% CI	
Parameter	Estimate	Std. Error	Df	t	Sig	Lower	Upper
Intercept	.8178	.2289	71.357	3.573	.001	.3615	1.2741
Location=1	.1352	.1910	35.489	.708	.484	-.2523	.5228
Location=2	.1523	.2164	44.926	.704	.485	-.2835	.5881
HDI=1	-.3665	.2751	79.401	-1.332	.187	-.9140	.1811
HDI=2	-.4914	.1625	60.944	-3.025	.004	-.8162	-.1665
Protein	-.0089	.0034	79.789	-2.640	.010	-.0156	-.0022
Location=1* Protein	-.0005	.0023	52.328	-.216	.830	-.0051	.0041
Location=2* Protein	-.0023	.0029	67.330	-.781	.438	-.0082	.0036
HDI=1 * Protein	.0049	.0037	85.441	1.325	.189	-.0024	.0122
HDI=2 * Protein	.0058	.0026	84.855	2.217	.029	.0006	.01094

Table (9) shows that: there is no significant interaction between "average protein consumption" in Asia"; But there is a significant interaction between "average protein consumption" in" Africa". However; intercepts and slope coefficient for the covariate differ according to "Location" and "HDI". Table (10) gives the predicted percent suffering from malnutrition according to Location and Human Development Index as function of average protein consumption.

Table (10): Prediction Equation According to Location and HDI

Location	**HDI**	**$\hat{y}$: Predicted Percent suffering frommalnutrition**
Asia	H	$\hat{y} = .586 - .005 \times \text{Protein}$
	M	$\hat{y} = .479 - .005 \times \text{Protein}$
	L	$\hat{y} = .953 - .009 \times \text{Protein}$
Africa	H	$\hat{y} = .603 - .006 \times \text{Protein}$
	M	$\hat{y} = .478 - .005 \times \text{Protein}$
	L	$\hat{y} = .970 - .001 \times \text{Protein}$
MENA	H	$\hat{y} = .451 - .004 \times \text{Protein}$
	M	$\hat{y} = .326 - .003 \times \text{Protein}$
	L	$\hat{y} = .818 - .009 \times \text{Protein}$

4.6 Comparison of Mixed Model to GLM Model

Obtained estimates using MLM are different than those obtained using GLM, especially that the design is unbalanced design. SPSS GLM procedure is used and the following table (Table 11) is obtained. The percent who suffers from malnutrion significantly different in the three human development index, However, the three "location" are not significantly different at the .05 level of significance i.e.; no effect for location on the dependent variable. There is significant interaction between location and the average grams consumed from protein, the percent of persons suffering from malnutrition is related to HDI, the error for the intercept is adjusted because the design is not balanced.

Table(11): Tests of Between-Subjects EffectDependent Variable: % Malnutrition

Parameter	Type III Sum of Squares	df	Mean Square	F	Sig.
Intercept:					
Hypothesis	.777	1	.777	31.130	.012
Error	.073	2.907	.025[b]		
HD					
Hypothesis	.168	2	.084[c]	7.633	.001
Error	.971	88	.011		
Location					
Hypothesis	.068	2	.084[c]	3.062	.052
Error	.971	88	.011		
Protein					
Hypothesis	.444	1	.444[c]	40.236	.000
Error	.971	88	.011		
Location* Protein					
Hypothesis	.103	2	.051[c]	4.647	.012
Error	.971	88	.011		

[b] .613(MS$_{Location}$)+.387(MS$_{Error}$)

[c] . MS$_{Error}$

GLM parameter estimates are exhibited in Table (12). Malnutrition in Asia is not affected by the average protein consumption; it is affected in Africa; also; Africa (Location 2) differs than the other two locations on the percent of persons suffering from malnutrition.

Table(12):GLM Parameter Estimates*

					95% CI	
Parameter	Estimate	Std. Error	t	sig	Lower	Upper
Intercept	.533	.135	3.940	.000	.265	.802
HDI=1	-.034	.050	-.684	.496	-.134	.065
HDI=2	-.112	.032	-3.465	.001	-.176	-.048
Location=1	.052	.158	.333	.740	-.261	.366
Location=2	.301	.150	2.009	.045	.003	.599
Protein	-.004	.002	-2.662	.009	-.005	-.001
Location=1* Protein	.000	.002	.152	.880	-.004	.004
Location=2* Protein	-.005	.002	-2.394	.019	-.009	-.001

*Dependent Variable: % Malnutrition

5. Conclusions

The Linear Mixed Models are preferred over GLM models when there are random effect factor, and when observations are clustered according to at least one factor, and when the design is not a balanced design. The Random Coefficient (RC) model with interaction is the OLS regression models with dummy variables and interaction with a covariate, however, the OLS assumes independence of observations, independent error term, and considers all effects as fixed. Thus the RC model with interaction is preferred when OLS assumptions are not satisfied. The study showed that the "average grams of protein intake" is the variable that affects the percentage of persons suffering from malnutrition, its estimate is -.0062 [Table 7], thus for each increase of 1 gram of protein/person, the percentage who suffers from malnutrition increase by approximately .01%; i.e., for each 10 grams increase, the decrease is approximately1%. However, the RC model with interaction [Table 9] reveals that the "protein intake" coefficients differ according to "Location" and HDI [Table 10].

References

I: Books and Journals

1. ADB (2001). Improving Children Nutrition in Asia. *Nutrition and Development Series* No. 3. Manila.
2. Atmarita, Jahari AB, Latief D, Soekirman, and Tilden RL (2000a). The Trend of Household Food Consumption Before and During Crises in Indonesia. *Proceedings of the 7th National Workshop on Food and Nutrition*, Jakarta, Indonesia.
3. Atmarita, Jahari AB, Latief D, Soekirman and Tilden RL (2000b). The Report of Assessments of School-entrance Children 1994 and 1999 Surveys", Directorate of Community Nutrition, Ministry of Health, Jakarta, Indonesia.
4. Dodge, Y. (2003).*"The Oxford Dictionary of Statistical Terms*", OUP, ISBN 0199206139
5. FAO. (1996). "The sixth world food survey". Rome. (1996).
6. FAO. (2003) "The State of Food Insecurity in the World 2003", Rome. (2003).
7. FAO. (2004). "The State of Food and Agriculture", Rome, (2003-04).
8. FAO. (2005). "*FAO* Statistical Yearbook 2005-2006".
9. Garson G.D. (2008). "Linear mixed models: random effects, hierarchical linear, multilevel, random coefficients, and repeated measures models (Statnotes style)", Lecturer notes at *Department of Public Administration, North Carolina State University,* pp. 1-49.
10. Hox, J. J. (1995). "Applied multi-level analysis",2nd Ed. Amsterdam: TT-Publicities. *A basic, non-technical introductory text.*
11. Hox, J. J. & Maas, C. J. M. (2001). "The accuracy of multilevel structural equation modeling with pseudo balanced groups and small samples". *Structural Equation Modeling* 8, 157-174.
12. James N. (2002a). "Malnutrition data, food security and the geography of food in communal area of North West Zimbabwe". *GBER* Vol. 2, No. 3, pp42-53.
13. James N. (2002b). "Behind Food Security: good Meals, new delicacies and crying for old foods. *Learning from food culture in Nembudzia*.
14. Kreft, I. G. G., de Leeuw, J., & Aiken, L. S. (1995). "The effect of different forms of centering in hierarchical linear models". *Multivariate Behavioral Research,* 30(1), 1-21.

15. Laird, N. M. & J. H. Ware (1982). “Random-Effects Models for Longitudinal Data.” Biometrics 38:963-974.
16. Moerbeek, M, Van Breukelen, GJP, & Berger, MPF (2008). "Optimal designs for multilevel studies". In: J de Leeuw & E Meijer (eds.). *Handbook of Multilevel Analysis*. New York: Springer.
17. Muthén, B.O. (1994). "Multilevel covariance structure analysis". *Sociological Methods and Research* 22: 376-399. .
18. Nyoman Latra, Susanti Linuwih, Purhadi2, and Suhartono (2010). "Estimation for Multivariate Linear Mixed Models", *International Journal of Basic & Applied Sciences* IJBAS-IJENS Vol.: 10 No: 06.
19. Searle S., Casella G. and McCulloch C. (2008). On line library, *John Wiley & Sons, Inc*., Hoboken, New Jersey.
20. Setboonsarng, S. 2005) "Child Malnutrition as a Poverty Indicator: An Evaluation in the Context of Different Development Interventions in Indonesia". *ADB Institute* Discussion Paper No. 21.
21. SPSS, Inc. "Linear mixed-effects modeling in SPSS: An introduction to the MIXED procedure". *SPSS Technical Report*. Chicago, IL: SPSS, Inc. (2005).
22. Stockholm International Water Institute *SIWI* (2005).
23. *United Nation, WFP*, "Local Estimation of Poverty and Malnutrition in Bangladesh", Bangladesh (2004).
24. *United Nations Development Program* (2005), United Nations.
25. Webb P. and k. Lapping (2002). "TUFTS NUTRITION", *The Gerald J. and Dorothy R. Friedman School of Nutrition Science and Policy*, FOOD POLICY AND APPLIED NUTRITION PROGRAM, Discussion Paper No.6[xix]
26. *World Bank* (2004). "Partnerships in Development-Progress in the Fight against Poverty", Washington D.C.

II. Internet Sites

---

[i] http://undp.org/en/media/HDR05_HDI1.pdf

[ii] *www.fao.org/economic/ess/publications.../fao...2005.../en/*

[iii] http://*malnutritionindevelopingcountries.blogspot.com*

[iv] http://who.int/water_sanitation_health/disease/malnutrition/en/

[v] *www.adbi.org/.../2005/...malnutrition.../measuring.malnutrition*

[vi] *www.who.int/healthinfo/statistics/bod_malnutrition.pdf*

[vii] (http://www.fao.org/unfao/govbodies/wfsfinal_en.htm).

[viii] Food Insecurity in the World
ftp://ftp.fao.org/docrep/fao/006/j0083e/j0083e00.pdf

[ix] Source: SIWI et al. 2005. Figure based on Tables 1 & 3 in FAO, 2004a)[ix]

[x] http://www.fao.org/unfao/govbodies/wfsfinal_en.htm, Cited1/7/2010 .

[xi] *www. ctl.scu.edu.tw/scutwebpub/website/.../91000128200932311474 8_2.pdf*

[xii] *www.psychology.unimaas.nl/.../Gerard_van_Breukelen_extended.htm*

[xiii] *www.faculty.chass.ncsu.edu/garson/PA765/multilevel.htm, cited on January 15, 2011*

[xiv] *www.projects.coe.uga.edu/jtemplin/files/glm2/psyc791s07_16.pdf*

[xv] http://projects.coe.uga.edu/jtemplin/files/glm2/psyc791s07_16.pdf, cited Jan. 20, 2011

[xvi] http://*onlinelibrary.wiley.com/doi/10.1111/j.1467-9469.2004.02...x/abstract*

[xvii] http://www.stat.sc.edu/scasa/Apr04t.html.

[xviii] *www.faculty.chass.ncsu.edu/garson/PA765/multilevel.htm, [cited on January 15, 2011].*

[xix] http://nutrition.tufus.edu/publications/ [cited Sep. 12. 2009]