DETERMINING THE RELATIONSHIP BETWEEN U.S. COUNTY-LEVEL ADULT OBESITY RATE AND MULTIPLE RISK FACTORS BY PLS REGRESSION AND SVM MODELING APPROACHES

Chau-Kuang Chen, Ed.D.
School of Graduate Studies and Research, Meharry Medical College
Nashville, TN 37208, U.S.A.

ABSTRACT

Data from the Center for Disease Control (CDC) has shown that the obesity rate doubled among adults within the past two decades. This upsurge was the result of changes in human behavior and environment.

Partial least squares (PLS) regression and support vector machine (SVM) models were conducted to determine the relationship between U.S. county-level adult obesity rate and multiple risk factors. The outcome variable was the adult obesity rate. The 23 risk factors were categorized into four domains of the social ecological model including biological/behavioral factor, socioeconomic status, food environment, and physical environment.

Of the 23 risk factors related to adult obesity, the top eight significant risk factors with high normalized importance were identified including physical inactivity, natural amenity, percent of households receiving SNAP benefits, and percent of all restaurants being fast food. The study results were consistent with those in the literature.

The study showed that adult obesity rate was influenced by biological/behavioral factor, socioeconomic status, food environment, and physical environment embedded in the social ecological theory. By analyzing multiple risk factors of obesity in the communities, may lead to the proposal of more comprehensive and integrated policies and intervention programs to solve the population-based problem.

Keywords: Adult Obesity, Food Environment, PLS Regression, SVM Model, and Social Ecological Theory,

INTRODUCTION

Obesity has become a significant public health concern in the United States. Data from the CDC showed that adult obesity rose in the 1990s ranging from 10 to 14 percent. In 2010, adult obesity ranged from 20 percent to 30 percent. Within the past two decades, the adult obesity rate increased by 10 to 16 percent [1]. If the obesity trend continues unabated, by 2030 approximately half of all U.S. adults will be obese [2]. The obesity epidemic is a major contributor to national morbidity and mortality [3, 4, 5, 6, and 7]. Public health prevention efforts are currently aimed at addressing this critical issue.

Various biological/behavioral factor, socioeconomic status, and food and physical environments among all U.S. counties may contribute to the complexity of this issue [8, 9, and 10].

Biological or Behavioral Factor

Researchers have shown that Blacks have the highest obesity rate (47.8%), followed by Hispanics (42.5%), Whites (32.6%), and Asians (10.8%) [1]. With respect to age, obesity was most prevalent amongst older men and middle-aged women [3].

Alcohol consumption has been commonly thought to increase the risk of obesity [11]. Drinking habits have consistently been linked with body weight. However, the use of alcohol actually lowered the risk of obesity compared to normal weight individuals [12].

Physical inactivity was associated with obesity and increased risk for chronic diseases (e.g., cardiovascular and certain cancers) and premature mortality [13]. Non-Hispanic Blacks were approximately 30 percent less likely to participate in physical exercise than any other ethnic group, regardless of income, education, and their awareness concerning a healthy diet or food choice [14]. This finding was a possible result of the unfavorable environment where most minorities typically reside [15].

Surrounding environments may be a limitation to people not participating frequently in physical exercise. With closer proximity to recreational facilities, shopping centers, and schools, physical activity has the potential to reduce the risk of obesity in many communities [16].

Socioeconomic Status

A study of BMI in relation to socioeconomic status was conducted [17]. The results indicated that lower levels of income and educational attainment usually result in BMI increases. Lower income people bought food based on price, taste, and convenience rather than healthy status [18]. Therefore, these choices contributed to the consumption of low-quality food, which ultimately lead to obesity.

The obesity rate was associated with percentage below poverty level, unemployment rate, and general income level. Although unemployment confined individuals to cheaper food options, employment also exaggerated increases in obesity [19]. By working long hours, employed individuals had easier access to unhealthy dieting due to time restraints.

Food Environment

Lower-priced food could lead to passive overeating due to energy-dense ingredients such as refined grains, added sugars, and fats [18]. Studies indicated that energy-dense foods and unhealthy eating habits contributed to poor nutrition and increased adult obesity rate.

Poor nutrition contributing to obesity rate was also linked to prices of foods [6, 20]. Food insecurity was a significant contributor to poor nutrition among adults [21]. Due to lack of income, individuals with such insecurity were more likely to purchase cheaper food products with low nutrition values.

The food environment consists of grocery stores, convenience stores, restaurants, and farmers' markets in public places [15].

Several studies revealed a strong relationship between healthy or unhealthy food environments and obesity rates [22, 23]. The ratio of convenience to grocery stores was selected as a risk factor due to the assumption that grocery stores typically offer more healthy foods compared to the unhealthy, energy dense processed foods primarily offered by convenience stores [24, 25].

Changes in the food environment have contributed to the increment of obesity rates from the early 1990s to the mid-1990s [26, 27]). During the same period, the 85 percent increase in fast-food restaurants corresponded with the 15 percent decrease of food stores [28]. The type of food availability in a community depends on neighborhood characteristics, such as race/ethnicity and socioeconomic status. Comparatively, supermarkets are more common among predominately white, high income communities while fast food restaurants and convenience stores are more common among minority, low income communities [29].

Fast food contains higher energy densities than any other foods. High energy dense food is usually harder for the digestive system to digest, therefore, increases the human body weight gain [30]. The availability of fast food restaurants resulted in weight gain while weight loss was evident where full-service restaurants were available. Therefore, a surrounding environment with multiple full-service restaurants was essential to healthy weight loss [24].

Physical Environment

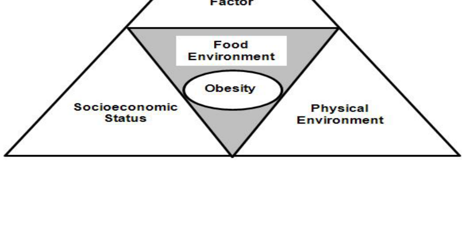
Neighborhood accommodations such as parks, sidewalks, protected bicycle lanes, and access to public transportation or recreational facilities usually lead to increased physical activity and ultimately, reduced adult obesity [28, 31]. The absence of these accommodations is expected to have long-term effects on residents in the neighborhood.

Additionally, the USDA's natural amenity index was included as a proxy for the influence of climate and the natural landscape (e.g., green space) on physical activity [10, 32]. Furthermore, the level of natural amenity (e.g., good or fair physical environments) in the geographical areas in the U.S. was significantly associated with a risk level (e.g., low or high) of obesity.

Application of Social Ecological Theory

Many health outcomes are functions of mass contributing risk factors [33, 34]. In planning for implementation of population-based prevention strategies, an approach integrating these different levels would greatly affect the lifestyle behaviors. As shown in Figure 1, the social ecological theory is used to determine the relationship between adult obesity and multiple risk factors encompassed within the environmental and the behavioral dominions.

Figure 1.
<u>Social Ecological Model</u>



It can be hypothesized that the obesity epidemic in all U.S. counties exhibits discernable patterns although adult obesity rate may be affected by biology/behavioral factors, socioeconomic status, and food and physical environments.

PLS Regression Model

The PLS regression model combines features of linear regression and principle component analyses. The great advantages of PLS regression over linear regression are (1) the availability of charts describing the model structure; (2) the capabilities of accommodating high-dimensional data and dealing with collinearity issues; and (3) not requiring the model assumptions such as normality, linearity, and independence.

The notion of PLS regression seeks to extract latent components, accounting for as much of the variations as possible in the model [35, 36]. Hence, the acronym PLS denotes "projection to latent structure". Therefore, PLS regression is capable of reducing a large number of variables into a few latent components (orthogonally spaced factors) that account for most of the variations in the dataset [37].

PLS regression simultaneously decomposes independent (X) and dependent (Y) variables to derive the orthogonal vectors (latent components) with the constraint that these components explain as much as possible of the covariance between X and Y matrices [38, 39, 40]). The covariance is a measure of how much X and Y variables affect each other. The two outer relation equations may be written as $X = TP' + E$ *and* $Y = UQ' + F$, where T and U are latent components, the extracted X- and Y-scores matrices, P and Q are coefficients, the X- and Y-loading matrices, and E and F are errors, the residual matrices [38, 39].

According to XLSTAT package, PLS regression allows researchers to minimize the residual matrices while maintaining the relationship between X and Y through the inner relation equation written as $U = TD$, where D is a diagonal matrix. Also, $T = XW^*$, where the latent components (T) are estimated as linear combinations of the original variables X_k with the weights W^*_{kl} (k = 1, 2, ..p; l = 1, 2, …, a, where a is the number of components)[41]. Based on the inner relation, the PLS regression can be rewritten as $Y = UQ' + F = XW^*C' + F = XB + F = XW(P'W)^{-1}C' + F$, with B ($p \times m$) and $B=W^*C'=W(P'W)^{-1}C'$, where weight vectors W and C (or W^*C values) [41].

SVM Model

The support vector machine (SVM) uses special non-linear functions called kernels to transform the input data into a multi-dimensional space [42]. The most suitable kernel function was chosen based on the model fitting statistic and overall prediction accuracy [43]. The SVM model seeks an optimal hyperplane to separate data from different categories through the computational shortcut of kernel functions [43]. The basic role of the kernel function is to calculate inner product values through a transformation in high-dimensional feature space, and ultimately maximize the margin of separation to yield high accuracy of data separation. The Decision Tree Regression (DTREG) software package was used to construct the SVM models [44]. All four kernel functions are involved in the SVM model construction: linear, radial basis function (RBF), polynomial, and sigmoid. Each SVM model contains only one kernel function to fit a hyperplane into the dataset [44].

Study Sample and Study Variables

The data points, 2,887 of all 3,144 U.S. counties in 2013, were analyzed. Approximately 8% (257) counties were excluded due to missing cells, (241) for multiple risk factors and outliers (16) for the initial run of PLS regression model. The outcome variable was the adult obesity rate defined as the percentage of adults in the county with a BMI $\geq 30kg/m^2$. The adult obesity rate in the United States was uniquely documented by data from the Behavioral Risk Factor Surveillance System [45]. As shown in Table 1, a total of 23 variables (risk factors) entered the PLS regression and SVM models via the XLSTAT and DTREG software packages. Data was compiled from multiple sources, including National Center for Chronic Disease Prevention and Health Promotion, the United States Census Bureau, and the United States Department of Agriculture (USDA) Economic Research Service Food Environment Atlas.

Table 1.
Study Variables

Variable		Measurement scales
Name	Description	
mdn_hh_income[1]	Median household income	Measured by dividing the income distribution into two equal groups: half has income above that amount and the other half has income below that amount.
unemploy_rate[1]	Unemployment rate	Measured by dividing the number of unemployed individuals by all individuals currently in the labor force.
poverty_rate[2]	Poverty rate	Measured annual poverty rates based on the summation of reported annual income reported a few times a year divided by the sum of poverty thresholds at month intervals.
pct_free_lunch[3]	Children eligible for free lunch program	Percent of children eligible for free lunch program.
pct_reduced_lunch[3]	Children eligible for reduced-price lunch program	Percent of children eligible for reduced-priced lunch program.
excess_drink[4]	Excessive drinking	More than 350,000 adults are interviewed each year using Behavioral Risk Factor Surveillance System (BRFSS) with a question “Have you had four or more drinks on one occasion?”
crime_rate[5]	Violent crime	Violent crime rate per 100,000 population.
physical_inact[6]	Physical inactivity	Percent of adults reporting no leisure-time physical activity.
access_rec_faci[7]	Access to recreational facilities (per 100,000 pop.)	Percent of people access to recreational facilities.
healthy_food[7]	Access to healthy food grocery stores	Percent of households with access to healthy food grocery stores.
Fast_food[7]	Fast food restaurants	Percent of all restaurants serving fast food.
rt_b_w_pop[8]	Black to white population ratio	Dividing the percentage of black population by the percentage of white population.
rt_ff_fs[8]	Ratio of fast-food to full-service restaurants	Dividing the percentage of fast-food restaurants /1,000 pop. by the percentage of full-service restaurants/1,000 pop.
rt_co_gr[8]	Ratio of convenience to grocery stores	Measured by dividing the percentage of convenience store/1,000 pop. by the percentage of grocery stores/1,000 pop.
pct_no_car_laccess[8]	Households with no car and low access to grocery stores	Percent of households with no car and low access to grocery stores
pct_snap[8]	Supplemental Nutrition Assistance Program (SNAP) participants	Percent of households that use SNAP
food_insecur[8]	Household food insecurity	Percentage of households with food insecurity
milk_usavg[8]	Price of low-fat milk, national average	Ratio of low-fat milk price, US national average
soda_price_usavg[8]	Price of soda, national average	Ratio of soda price, US national average
milk_soda_price[8]	Price of low-fat milk/soda, national average	Divide the national average of low-fat milk price by the national
natural_amenity[8]	USDA’s natural amenity index	Measuring each county by a scale with the score of 3 being the most desirable county to live, and over -2 as the least desirable county to live.
Perpovcnty[8]	Persistent Poverty County	Economic Resource Service (ERS) has defined counties as being persistently poor if 20 percent or more of their
Metro[8]	Metropolitan or Non-metropolitan Area	Metropolitan areas have a 50,000 or more pop. while Non-metropolitan areas have less than 50,000 pop.

[1]United States Census Bureau
[2]Office of Management and Budget’s Statistical Policy Directive 14
[3]National Center for Education Statistics

[4]Center for Disease Control and Prevention
[5]Uniform Crime Reporting – FBI
[6]National Center for Chronic Disease Prevention and Health Promotion
[7]County Business Patterns
[8]United States Department of Agriculture Economic Research Service Food Environment Atlas

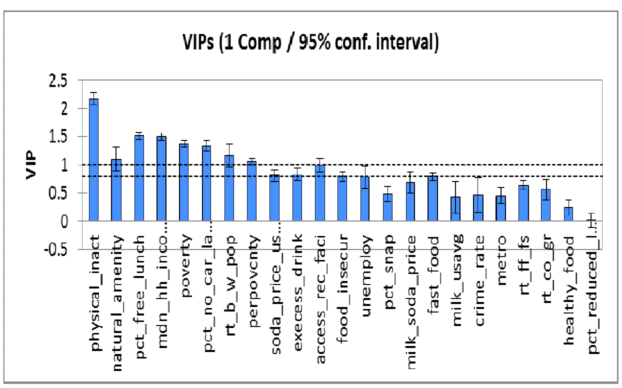
STUDY METHOD

A graphical method for variable selection is available in XLSTAT software, which is the quantity of Variable Importance in the Projection (VIP). Variables with VIP values close to one can be considered important in the model. Additionally, the biplot graph enables researchers to visualize the commonality (possession of common features) among risk factors. In particular, the higher one risk factor is correlated to the others, the closer in proximity the risk factor is to each other on the graph. Furthermore, a graph of 95% confidence intervals for standardized coefficients is constructed. It displays the directions and strengths of the relationship between outcome variable and its risk factors (individual risk factor is significant if the interval does not contain the null (zero) value). Finaly, data analysis was carried out by comparing the normalized importance, rank order of risk factors, and prediction accuracy on the PLS regression to those of SVM model.

STUDY RESULTS

As shown in Figure 2, the top eight important risk factors that contributed to the adult obesity rate were: (1) percent of adults reporting no leisure-time physical activity (physical_inact); (2) natural amenity (natural enmity); (3) percent of children eligible for free lunch program (pct_free_lunch); (4) median household income (mdn_hh_income); (5) poverty rate (poverty_rate); (6) percent of households with no car and low access to grocery stores (pct_no_car_laccess); (7) black to white population ratio (rt_b_w_pop), and (8) persistent poverty county (perpovcnty) on component 1.

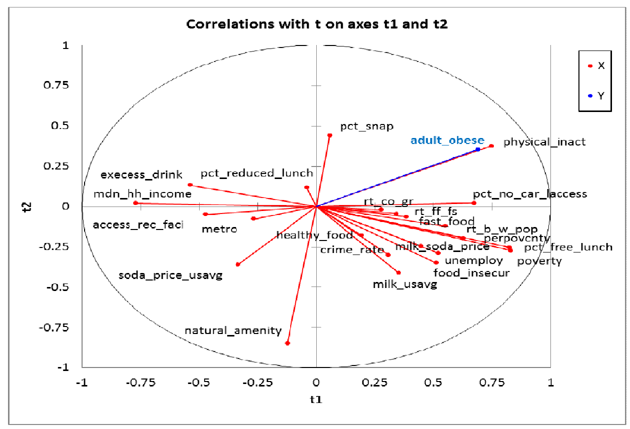
Figure 2.
<u>VIP Values for Component 1</u>



A biplot in Figure 3 indicated that the relationships between latent components and adult obesity rate demonstrate the unique feature of PLS regression. Four risk factors related to food environment were closely aligned with the latent component (t2) and, in fact, were in the same positive direction as the adult obesity rate. These four risk factors were collapsed into the latent component (t2), namely food environment that was positively related to the adult obesity rate as follows: (1) percent of households with no car and low access to grocery stores (pct_no_car_laccess); (2) ratio of convenience to grocery stores (rt_co_gr), (3) ratio of fast-food to full-service restaurants (rt_ff_fs), and (4) percent of all restaurants serving fast food (fast_food). However, the following two risk factors in the negative direction of the latent component (t2), namely socioeconomic status that negatively contributed to the adult obesity rate: (1) median household income (mdn_hh_income); and (2) percent of people access to recreational facilities (access_rec_faci).

Overall, the findings just mentioned above were consistent with those in Figure 2, where top eight important risk factors were significantly associated with the adult obesity rate.

Figure 3.
<u>Risk Factors Correlated with Latent Components (t1 and t2)</u>



<u>Model Fitting and Accuracy</u>

Both PLS and SVM models were found to be good fit with R squared values of 0.642 and 0.692, indicating that all risk factors entering this model accounted for approximately 64% and 69% of the variation in adult obesity rate. In addition, PLS and SVM models had small mean absolute percent error (MAPE) values of 2.72% for PLS and 6.12% for SVM, indicating that the models had high percent of predictive accuracy.

MAJOR FINDINGS

As shown in Table 2, the eight significant and normalized important risk factors related to adult obesity were selected from the following three criteria generated from both models: (1) top eight ranking with a VIP value greater than 1 in PLS regression; (2) top eight ranking in the absolute value of the PLS standardized coefficent that must be statistically significant; and (3) top eight ranking of normalized importance in SVM model.

Of all eight significant risk factors with high normalized importance, three risk factors meeting all afoermentioned criteria were: (1) percent of adults reporting no leisure-time physical activity, (2) natural amenity, and (3) black to white population ratio. Also, the study showed that two risk factors meeting the first two criteria: (1) children eligible for free lunch program and (2) percent of households with no car and low access to grocery stores. Additionally, median household income met the first and third criteria while the percent of households receiving SNAP benefits and percent of all restaurants being fast food met the last two criteria.

Eight risk factors belonged to the four dominions in the social ecological theory. These risk factors included: (1) Biological or behavioral factor--black to white population ratio and percent of adults reporting no leisure-time physical activity; (2) Socioeconomic status--median household income, percent of children eligible for free lunch program, and percent of households receiving SNAP benefits; (3) Food environment--percent of all restaurants serving fast food, and percent of households with no car and low access to grocery stores; and (4) Physical environment--natural amenity index.

Table 2.
Significant and Normalized Important Risk Factors in PLS Regression and SVM Models

Risk factors in equation	PLS regression		SVM*	Sig. and/or important risk factors to both PLS and SVM
	Top 8 VIP with value > 1	Top 8 absolute values of sig. std coeff	Top 8 NI	
physical_ inact	1	1	1	Yes
natural_ amenity	2	2	2	Yes
pct_free_ lunch	3	4		Yes
mdn_hh_ income	4		6	Yes
pct_no_car_ laccess	6	8		Yes
rt_b_w_pop	7	3	3	Yes
pct_snap		5	5	Yes
fast_food		6	8	Yes

[a] The best kernel function in SVM model was RFB rather than linear, polynomial, and sigmoid, which yielded the smallest MAPE value of 6.12% with parameter values: Epsilon = 0.001, C = 40.7886, Gamma = 0.4079, and P = 1.

CONCLUSIONS AND IMPLICATIONS

This study accomplished its objective by confirming the research hypothesis that the obesity epidemic in all U.S. counties exhibits discernable patterns affected by four domains of social ecological framework: biological or behavioral factor, socioeconomic status, food environment, and physical environment. Top eight significant and normalized important risk factors such as percent of adults reporting no leisure-time physical activity and percent of children eligible for free lunch program were embedded in these domains, demonstrating that PLS regression and SVM models yield the construct validity as well as the predictive validity.

In terms of the biological or behavioral factor, black to white population ratio and percent of adults reporting no leisure-time physical activity, were in line with the findings of adult obesity studies in the literature [46]. Also, three significant and normalized important risk factors were clustered together as a latent component, namely socioeconomic status, correlated with adult obesity rate: median household income, percent of children eligible for free lunch program, and percent of households receiving SNAP benefits, which were also in consistent with the adult obesity literature [15]. Additionally, two significant and normalized important risk factors found in the food environment congruent with the literature [24] were: percent of all restaurants serving fast food and percent of households with no car and low access to grocery stores. Furthermore, for the physical environment in this study, the natural amenity affecting adult obesity was in accordance with the obesity literature [10].

The principle of social ecological model regarding adult obesity proposes that creating an environment conducive to change is important to sustain healthy behaviors. The full integration of these population-based interventions that span multiple domains could lead to a significant decrease in the adult obesity rate. Population-based interventions should be conducted in schools, community centers, and churches to focus on human behavior modification in order to engage in physical activity.

REFERENCES

[1] Obesity and overweight for professionals: Data and statistics, **CDC,** 2012, Retrieved from http://www.cdc.gov/obesity/data/ adult.html. August 13, 2014.

[2] Y.C. Wang, K. McPherson, T. Marsh, S.L. Gortmaker, and M. Brown, Health and economic burden of the projected obesity trends in the USA and the UK, **Lancet,** 2011, 378:815-25.

[3] K.M. Flegal, M.D. Carroll, C.L. Ogden, and L.R. Curtin, Prevalence and trends in obesity among US adults, 1999-2008, **JAMA,** 2010 Jan 20;303(3):235-41

[4] N.S. Wellman and B. Friedberg, Causes and consequences of adult obesity: health, social and economic impacts in the United States, **Asia Pac J Clin Nutr,** 2002, *11 Suppl 8*, S705-709.

[5] National Institutes of Health, Clinical Guidelines on the Identification, Evaluation, and Treatment of Overweight and Obesity in Adults--The Evidence Report, **Obes. Res.,** 1998, Sep;6 Suppl 2:51S-209S.

[6] S. Kumanyika, et al., Population-Based Prevention of Obesity: The Need for Comprehensive Promotion of Healthful Eating, Physical Activity, and Energy Balance. **AHA Scientific Statement,** 2008, 118:428-464.

[7] R.M. Benjamin, The Surgeon General's vision for a healthy and fit nation, **Public health reports,** 2010, *125*(4), 514.

[8] T. Batts, C.K. Chen, J. Buford, and V. Brown, Investigating Risk Factors Associated with Obesity Rates in the United States based on PLS Regression and ANN Models, **European International Journal of Science and Technology,** 2014, 3 (1), 43-58.

[9] N. Balthasar, L.T. Dalgaard, C.E. Lee, J. Yu, H. Funahashi, T. Williams, and B.B. Lowell, Divergence of melanocortin pathways in the control of food intake and energy expenditure, **Cell,** 2005, *123*(3), 493-505.

[10] S.H. Chi, D.S. Grigsby-Toussaint, N. Bradford, and J. Choi, Can geographically weighted regression improve our contextual understanding of obesity in the US? Findings from the USDA Food Atlas, **Applied Geography,** 2013, 44, 134-142.

[11] Davis, S. Facts on alcohol and obesity, Retrieved from http://www.livestrong.com/article/69121-alcohol obesity/, 2010.

[12] A.N. Gearhardt, E.L. Harrison, and S.A McKee, Does co-morbid depression alter the inverse relationship

between obesity and substance use disorders?, **Drug and alcohol dependence**, 2012, *124*(1), 185-188.

[13] Physical activity and health: a report of Surgeon General, Atlanta, Georgia: U.S. Department of Health and Human Services, **CDC,** 1996.

[14] Y. Wang and X. Chen, How much of racial/ethnic disparities in dietary intakes, exercise, and weight status can be explained by nutrition- and health-related psychosocial factors and socioeconomic status among US adults, **Journal of the American Dietetic Association,** 2011, *111*, 1904-11.

[15] J.F. Sallis and K. Glanz, Physical activity and food environments: solutions to the obesity epidemic. **Milbank Quarterly,** 2009, 87(1), 123-154.

[16] L.D. Frank, M.A. Amdresen, and T.L. Schmid, Obesity relationships with community design, physical activity, and time spent in cars, **Am. J Prev. Med.,** 2004. 27(2):87-96.

[17] M. Grabner, BMI trends, socioeconomic status, and the choice of dataset, **Obesity Facts,** 2012, *5*(1), 112-126.

[18] A. Drewnowski and N. Darmon, The economics of obesity: dietary energy density and density cost, **Am. J Clin. Nutr.,** 2005, 82(suppl): 265S-273S.

[19] N.F Weaver, L. Hayes, N.C. Unwin, and M.J. Murtagh, "Obesity' and 'Clinical Obesity": Men's understandings of obesity and its relations to the risk of diabetes: A qualitative study, **BioMedCentral Public Health,** 2008, *8*(311), 1-8.

[20] M. Simon, Follow the Money FOOD Profiting from. **Eat, Drink, Politics,** 2012, Available at: http://www.eatdrinkpolitics.com/wp-content/uploads/FoodStampsFollowtheMoneySimon.pdf

[21] J. Bhattacharyaa, J. Currie, and S. Haiderc, Poverty, food insecurity, and nutritional outcomes in children and adults, **Journal of Health Economics,** 2004, 23, 839–862.

[22] J.N. Bodor, J.C. Rice, T.A Farley, C.M. Swalm, and D. Rose, The association between obesity and urban food environments, **Journal of Urban Health,** 2010, 87(5), 771-781.

[23] J.C. Spence, N. Cutumisu, J. Edwards, K.D. Raine and K. Smoyer-Tomic, Relation Between Local Food Environments and Obesity Among Adults. **BMC Public Health**, 2009, 9:192 doi:10.1186/1471-2458-9-192.

[24] N.K. Mehta and V.W. Chang, Weight loss and restaurant availability: A multilevel analysis, **Am. J Prev. Med.** , 2008, 34(2):127-33.

[25] A. Michimi and M. C. Wimberly, Associations of supermarket accessibility with obesity and fruit and vegetable consumption in the conterminous United States, **International Journal of Health Geographics,** 2010, 9(49), 94-112.

[26] S.A. Bowman and B.T. Vinyard, Fast food consumption of US adults: impact on energy and nutrient intakes and overweight status. **Journal of the American College of Nutrition,** 2004, *23*(2), 163-168.

[27] Obesity trends among U.S. adults between 1985 and 2010, Centers for Disease Control and Prevention, 2010, Retrieved from: http://www.cdc.gov/obesity/ downloads/ obesity_trends_2010.pdf.

[28] D.A. Cohen, Obesity and the built environment: changes in environmental cues cause energy imbalances, **International Journal of Obesity, 2008**, 32, S137-S142.

[29] L.V. Moore and A.V. Diez Roux, Associations of neighborhood characteristics with the location and type of food stores, Am **J Public Health,** 2006, 96:325–31.

[30] A.M Prentice and S.A. Jebb, Fast foods, energy density and obesity: A possible mechanistic link. **Int. Assn. Obesity Reviews**, 2003, 4:187-94.

[31] Transportation Research Board and Institute of Medicine, Does the Built Environment Influence Physical Activity? **Examining the Evidence, Special Report 282**, 2005, Washington, DC: National Academies Press.

[32] P. Tucker and J. Gilliland, The effect of season and weather on physical activity: a systematic review, **Public Health,** 2007 121(12), 909-922.

[33] L.M. Carrascal, I. Galván, and O. Gordo, Partial least squares regression as an alternative to current regression methods used in ecology. **Oikos,** 2009;118(5):681–690. doi:10.1111/j. 1600-0706. 2008.16881.x.

[34] N. Sellin, Partial least squares analysis, ***International Journal of Educational Research*****,** 1986, 10(2), 189-200.

[35] B.D. Fekedulegn, J.J Colbert, R.R. Hicks, and M Schuckers, Coping with Multicollinearity: An Example on Application of Principal Components Regression in Dendroecology, 2002, United States Department of Agriculture.

[36] D.T. Tobias, An Introduction to Partial Least Squares Regression, **Analysis,** 2003(16).

[37] C.N. Alves, J.C. Pinheiro, A.J. Camargo, M.M.C. Ferreira, R.A.F. Romero, and A.B.F. da Silva, A multiple linear regression and partial least squares study of flavonoid compounds with anti-HIV activity, **Journal of Molecular Structure,** 2001 (541), 81-88.

[38] Abdi Partial Least Squares. In: M. Lewis-Beck, A. Bryman, and T. Futing, eds., **Encyclopedia of Social Sciences Research Methods,** Thousand Oaks, CA: Sage; 2003,792–5.

[39] R. Rosipal and N. Kramer, Overview and Recent Advances in Partial Least Squares. In: Saunders C, ed. **Lecture Notes in Computer Science**, 2006, Berlin Heidelberg: Springer-Verlag; 34–51. Available at: http://link.springer.com/ chapter/10.1007/ 11752790_2#.

[40] S. Maitra and J. Yan, Principle Component Analysis and Partial Least Squares: Two Dimension Reduction Techniques for Regression, **Casualty Actuarial Society**, 2008 Discussion Paper Program.

[41] Ardinsoft Company Running a Partial Least Squares regression with XLSTAT, 2012

[42] V. Vapnik, **The nature of statistical learning theory,** 1995, New York: Springer-Verlag.

[43] N. Cristianini and J. Shawe-Taylor, An introduction to support vector machine and other kernel- based learning methods. Cambridge, England: Cambridge University Press, 2000.

[44] P.H. Sherrod, *DTREG predictive modeling and forecasting*, 2007, Retrieved November 21, 2010, from http://www.dtreg.com.

[45] E.S. Ford and A.H., Epidemiology of Obesity in the Western Hemisphere, **J Clin Endocrinol Metab.,** 2008 Nov;93(11 Suppl 1):S1-8. doi: 10.1210/jc.2008-1356.

[46] M. Wen, and L. Kowaleski-Jones, Sex and ethnic differences in validity of self-reported adult height, weight and body mass index, **Ethnicity & Disease,** 2012, *22*(1), 72.