PJN
ISSN 1680-5194

PAKISTAN JOURNAL OF NUTRITION

ANSInet
308 Lasani Town, Sargodha Road, Faisalabad - Pakistan
Mob: +92 300 3008585, Fax: +92 41 8815544
E-mail: editorpjn@gmail.com

Pakistan Journal of Nutrition 9 (6): 611-615, 2010
ISSN 1680-5194

Median Regression Analysis of Body Mass Index of Adults in Pakistan

Muhammad Aslam[1], Aamir Saeed[2], G.R. Pasha[3] and Saima Altaf[4]
[1]Department of Statistics, Bahauddin Zakariya University, Multan, Pakistan
[2]District Data Manager (Health), NCHD, Bahawalpur, Pakistan
[3]BZU Sub-campus, D.G. Khan, Pakistan
[4]Department of Statistics, PMAS Arid Agricultural University, Rawalpindi, Pakistan

Abstract: Body Mass Index (BMI) is considered to be the most popular measure for overweight and obesity. Numerous studies of BMI are limited to compute and interpret different percentiles of BMI and do not account for many other covariates affecting BMI. Conventional regression methods are used for estimating how covariates are related to mean values of the dependent variable but in many situations, we are interested in quantiles rather than in mean values as in the case of BMI analysis. The present study addresses the same using median regression. Some important covariates such as gender, age, marital status, daily working hours, daily exercise routine and number of meat-eaten days per week are included in the study and found to be significant.

Key words: Obesity, quantile, quantile regression

INTRODUCTION

Overweight is a result of an imbalance between energy intake and expenditure while the term, obesity is used to describe body weight that is much greater than what is considered healthy. It is well established that obesity is associated with adverse health effects, e.g., gall bladder disease, hypertension, sleep apnea, gout, breast and endometrial cancer, colorectal cancer and osteoarthritis etc. (Bray *et al*., 1998; Marion and Jacobson, 2000; Ferris, 2007). There are approximately 350 million obese people and over 1 billion overweight people in the world. Over all about 2.5 millions deaths are attributed to overweight/obesity worldwide (Siervo *et al*., 2007). In developed countries, obesity is one of the aggravated public health problems (Zohoori *et al*., 1998; Mokdad *et al*., 2002; Peytremann-Bridevaux, 2007). In the US alone, it was estimated (see Golditz, 1999) that excess weight and physical inactivity accounted for 300,000 premature deaths per year and for 9.4% of all direct health care costs ($70 billion attributable to obesity). According to Bovet *et al*. (2004), the prevalence of overweight has increased greatly in developed countries over the last two to three decades in adults, children and infants. Although few data are available in developing countries, the epidemic of obesity is also occurring in these countries and Pakistan is one of them. Many researchers have focused their interest to study the overweight/obesity prevalence in Pakistan, see for example, Bharmal, 2000, Pappas *et al*., 2001; Nanan, 2002; Afridi and Khan, 2004 etc. and recently, Aslam *et al*. (2010).

Usually, weight status is determined by a person's BMI, defined as the ratio of weight (kg) to squared height in meters (m^2). According to WHO's standards, a person is overweight if BMI $\geq$ 25 and is obese if BMI $\geq$ 30. Unlike the many earlier studies in Pakistan (Kiyani *et al*., 2002; Rehman *et al*., 2003; Shah *et al*., 2004; Khan *et al*., 2008) on BMI, Aslam *et al*. (2010) use the new recommendation of WHO (2000) for Asia Pacific Region. According to this recommendation, in Pakistan, a person will be underweight (if BMI $\leq$ 19), normal (if 20 $\leq$ BMI < 23), overweight (if 23 $\leq$ BMI < 25) and obese (BMI $\geq$ 25) (see, also Nanan, 2002; Leung *et al*., 2008; Jaleel, 2009). Aslam *et al*. (2010) report that more than 46% people are overweight or obese in Multan. The common point in all the studies, discussed above, is that all of them rely on just computing BMI and percentiles. They do not take into account different covariates, responsible for obesity. According to Kan and Tsai (1993), the possession of knowledge on obesity's health risks prevents an individual from being overweight so it is useful to see the impact of different responsible covariates on BMI. This fact motivates the present article and thus, it is a continuation of the work done by Aslam *et al*. (2010). Following Chen (2005) and Ruhm (2006), we make use of median regression, a special case of quantile regression, for studying the effects of different covariates on BMI.

Median regression: According to Chen (2005), the percentiles of BMI for a specified age are of particular interest in light of public health concerns. The empirical percentiles with grouped age provide a discrete approximation for the population percentile so it is more plausible to employ some regression methods to study the effects of different factors on obesity prevalence. He adds further that for the obesity prevalence rates with the relation to different factors, usual regression methods

Corresponding Author: Dr. Muhammad Aslam, Department of Statistics, Bahauddin Zakariya University, Multan, Pakistan

do not fulfill the desired objectives. Traditional regression methods are useful for estimating how covariates are related to mean values of the dependent variable but, without strong parametric assumptions, will not accurately indicate changes at other points in the distribution. Such assumptions are unlikely to be justified since BMI increases more over time at the high than the low end of the distribution. Kan and Tsai (1993) also report, in their study, that there is a steeper increase in the BMI towards the right tail of the distribution. As an alternative, quantile regression methods are used for such type of analysis. Quantile regression, proposed by Koenker and Bassett (1978), minimizes the weighted sum of the absolute deviations of the error term, unlike regression models that minimize the sum of the squared residuals. According to Koenker and Hallock (2001), the purpose of the quantile regression is to estimate conditional quantile functions, where quantiles of a response variable's distribution are specified as functions of observed covariates (see also, Flegal and Troiano, 2000).

To briefly recall the ordinary quantile, consider a real valued random variable Y characterized by the following distribution function,

$$F(y) = \text{Prob}\,(Y \le y)$$

The τ-th quantile of Y is defined as the inverse function

$$Q(\tau) = \inf\{y: f(y) \ge \tau\}$$

Where $0 < \tau < 1$. In particular, the median is $Q(1/2)$.

The τ-th sample quantile $\hat{\xi}(\tau)$, which is an analogue of $Q(\tau)$, may be formulated as the solution of the optimization problem,

$$\min_{\xi = R} \sum_{i=1}^{n} \rho_\tau (y_i - \xi)$$

Where $\rho_\tau(z) = z[\tau - I(z < 0)]$, $0 < \tau < 1$, is usually called the check function.

When covariates X (e.g., age, gender etc.) are considered, the linear conditional quantile function, $Q(\tau | X = x) = x'\beta(\tau)$, can be estimated by solving,

$$\hat{\beta}(\tau) = \text{argmin} \sum_{i=1}^{n} \rho_\tau (y_i - x'_i \beta) \qquad (1)$$

for any $\tau \in (0,1)$. The quantity $\hat{\beta}(\tau)$ is called the regression quantile. The case $\tau = 1/2$, which minimizes the sum of absolute residual, is usually known as median regression. For more details about median regression, see Koenker and Hallock (2001), Buhai (2004), Martins and Pereira (2004) and Chen and Wei (2005).

MATERIALS AND METHODS

A cross-sectional data comprising of 2000 adult (aged 14 years or more) individuals, both males and females were taken from Multan city from January 1, 2007 to December 31, 2008 as a case study of Pakistan (see, Aslam *et al*., 2010 for more details about the data).

For the present study, following Gortmaker *et al*. (1993), Chen (2005) and Ruhm (2006), we take data on different variables, including the factors responsible for obesity determination. These variables, with respective codes and values shown in parentheses, are Gender (GENDER: 1 = male and 2 = female), Age in years (AGE, rounded to next year), Marital Status (MSTAT: 0 = single, 1 = married), Weight in Kg. (WT), Height in inches (HT), Hours Worked in Field per day (FWH: No. of daily working hours in field), Hours Worked at Home per day (HWH: No. of daily working hours at home), Daily Exercise Routine (EXER: 0 = no exercise, 1 = irregular exercise, 2 = regular exercise) and No. of Meat-eaten Days per Week (MWK: No. of days per week when any type of meat is taken in the meal of the respondent).

BMI of the individuals are calculated as weight in kilograms divided by height in meters squared. BMI is taken as dependent variable (y_i) and vector of regressors, X_i including AGE, MSTAT, FWH, HWH, EXER, and MWK for median regression equation, defined in (1). Since men and women have different growth patterns (Chen, 2005), median regression analysis of BMI is split into two sets, one for males and other for females. Chen (2005) uses SAS QUANTREG procedure but we use the software package, STATA 10.0 for the computations.

RESULTS AND DISCUSSION

In our data set of 2000 individuals, 1123 are males (56.2%) and 877 are females (43.8%). The details of summary statistics about the age, marital status, weight, height, BMI, BMI percentiles and obesity status can be had from Aslam *et al*. (2010) while the same about the rest of variables is presented here.

Table 1 reports the first quartile (Q1), median, third quartile (Q3), mean and standard deviation (SD) for age, weight, height and BMI of both the genders. The empirical value of median for BMI will be used next to compare with the estimates of median regression.

Table 2 shows daily working hours of the respondents working in field and at home. This table reflects the physical exertion of the respondents when working. However, the table does not reflect the nature of the work and is not gender discriminating as males are expected to do more in field as compared to do at homes and for

Table 1: Summary statistics

Variable	Q1	Median	Q3	Mean	SD
Males					
Age	20.00	24.00	30.00	25.92	7.61
Weight	62.00	68.00	74.00	68.11	10.64
Height	65.00	67.00	69.00	67.06	3.02
BMI	20.91	23.48	25.81	23.51	3.63
Females					
Age	19.00	21.00	23.00	22.33	5.41
Weight	48.40	54.00	61.00	55.23	9.69
Height	61.00	62.00	64.00	62.38	2.63
BMI	19.33	21.53	24.27	22.05	3.92

Table 2: Daily working hours

Working hours	In field		At home	
	Frequency	%age	Frequency	%age
<4	357	17.85	1593	79.65
5-9	1193	59.65	339	16.95
10-14	438	21.90	67	03.35
15 and above	12	00.60	01	00.05
Total	2000	100.00	2000	100.00

females, this fact is vice versa. It is noted that majority of the respondents (59.65%), work for 5-9 hour daily in field while only 0.60% do work for 16 or more hours daily in field. On the other hand, about 80% of the respondents do work for less than 4 hour at home.

The daily exercise routine of the respondents is given in Table 3. The table shows that majority of the respondents (about 70%) do not take any exercise while just 12% take some regular exercise.

Table 4 shows the no. of days in which the respondents take meat in their meals per week. It can be noted that only 1.2% of the respondents do not take meat in their foods at all while about 20% take meat daily. It is also reported that about 43% of the respondents take meat in 3 or less days of a week. These figures, however, do not give any information about the quantity, type and form of the meat taken.

Table 3: Daily exercise routine

Daily exercise	Frequency	%age
No exercise	1394	69.7
Irregular exercise	367	18.4
Regular exercise	239	12.0
Total	2000	100.0

Table 4: No. of meat-eaten days per week

No. of days	Frequency	%age	Cumulative %age
0	24	01.2	01.2
1	109	05.5	06.7
2	305	15.3	21.9
3	426	21.3	43.2
4	378	18.9	62.1
5	214	10.7	72.8
6	153	07.7	80.5
7	391	19.6	100.0
Total	2000	100.0	

The parameter estimates of median regression for males and females are given in Table 5 and 6, respectively. It is noted that majority of the coefficients are statistically significant at 1% level of significance except EXER (at 5%) for females while MSTAT, FWH and EXER are significant at 5% level of significance for males. By using these estimates, one can easily estimate 50th percentile of BMI for any adult male or female. For illustration purpose, suppose one considers an unmarried male (MSTAT = 0) of age 22 years who works for 6 hour in field daily (FWH = 6) and for 1 hour at home (HWH = 1), does not take any regular exercise (EXER = 0) and eats meat four days in a week (MWK = 4). Using Table 5, the 50th percentile of BMI for such kind of males, is 22.49. Thus, median regression estimates the 50th percentile (i.e., median) to be 22.49 by incorporating all the above stated significant covariates while the empirical median of BMI for all ages is 24 (Table 1). A similar practice can readily be done to compute 50th percentile for females using Table 6.

Table 5 : Parameter estimates with median regression ($\tau = 1/2$) for males

	Coeff.	Std. Error	t	P > t	95% Confidence Interval	
Constant	17.52	0.724	24.19	0.0000	16.10	18.94
MSTAT	0.81	0.409	1.99	0.0470	0.01	1.62
AGE	0.16	0.025	6.52	0.0000	0.11	0.21
FWH	0.01	0.005	2.27	0.0234	0.00	0.02
HWH	-0.22	0.060	-3.59	0.0000	-0.34	-0.10
EXER	-0.06	0.197	2.05	0.0402	-0.45	0.32
MWK	0.40	0.071	5.61	0.0000	0.26	0.54

Table 6 : Parameter estimates with median regression ($\tau = 1/2$) for females

	Coeff.	Std. Error	t	P > t	95% Confidence Interval	
Constant	14.85	0.811	18.32	0.0000	13.26	16.45
MSTAT	2.67	0.465	5.74	0.0000	1.76	3.58
AGE	0.16	0.031	5.02	0.0000	0.09	0.22
FWH	0.19	0.049	3.95	0.0000	0.10	0.29
HWH	0.23	0.054	4.25	0.0000	0.12	0.34
EXER	-0.22	0.110	-2.04	0.0416	-0.44	-0.01
MWK	0.29	0.075	3.82	0.0000	0.14	0.43

Conclusion: Obesity is considered to be epidemic health problem worldwide. Being overweight is recognized as a significant risk factor for numerous diseases. Consequently, the prevalence of obesity needs to be observed, seriously. Unlike usual studies, the present work focuses on evaluating the dependence of BMI on many significant covariates including age, marital status, physical exertion reflected by daily working hours in field and at homes, routine about daily exercise and meat intake per week. The results show that more than 60% people work in field for 5 or more hours and a majority (about 80%) does less than 4 hour at homes. Majority of the respondents does not take any regular exercise. It is also reported that only 1.2% of the respondents do not take any meat in their foods while about 20% take meat daily. The median regression analysis shows that all the above stated covariates play a significant role in the prevalence of obesity. These estimates are obtained for both, males and females, separately and one can find 50^{th} percentile of BMI for males and females using these estimates against given covariates. With the help of quartile regression, we can estimate other percentiles choosing appropriate τ ($0 < \tau < 1$).

REFERENCES

Afridi, A.K. and A. Khan, 2004. Prevalence and etiology of obesity - an overview. Pak. J. Nutr., 3: 14-25.

Aslam, M., A. Saeed, G.R. Pasha and S. Altaf, 2010. Gender differences of body mass index in adults of Pakistan: A case study of Multan city. Pak. J. Nutr., 9: 162-166.

Bharmal, F.Y., 2000. Trends in nutrition transitions: Pakistan in focus. J. Pak. Med. Assoc., 50: 159-167.

Bovet, P., G. Madeleine, D. Padayachy and N. Stettler, 2004. Increasing prevalence of overweight among Seychelles children, 1998-2002. MDJ Seychelles Medical and Dental J., Vol. 7.

Bray, G.A., C. Bouchard and W.P.T. James, 1998. Handbook of Obesity, New York.

Buhai, S., 2004. Quantile regression: Overview and selected applications. Tinbergen Institute and Erasmus University, Rotterdam.

Chen, C., 2005. Growth charts of Body Mass Index (BMI) with quantile regression. SAS Institute Inc. Cary, NC, USA.

Chen, C. and Y. Wei, 2005. Computational issues on quantile regression, Special Issue on Quantile Regression and Related Methods. Sankhya.

Ferris, M., 2007. Obesity, albuminuria and urinalysis findings in US young adults from the add health wave iii study. Clin. J. Am. Soc. Nephrol, 2: 1207-14.

Flegal, K.M. and R.P. Troiano, 2000. Changes in the distribution of body mass index of adults and children in the US population. Int. J. Obes., 24: 807-818.

Golditz, G.A., 1999. Economic costs of obesity and inactivity. Med. Sci. Sports Exerc, 31: S663-667.

Gortmaker, S.L., A. Must and J.M. Perrin, 1993. Social and economic consequences of overweight in adolescent and young adulthood. N. Eng. J. Med., 329: 1008-1012.

Jaleel, R., 2009. Impact of maternal obesity on pregnancy outcome. J. Surgery Pak. (International), Vol. 14.

Kan, K. and W. Tsai, 1993. Obesity and risk knowledge in Taiwan: A quantile regression analysis. JEL Classification: I12, I18, C51.

Khan, M.H., H. Khan, G. Sarwar, B. Iftikhar, M. Naimatullah and A. Gul, 2008. Study of obese persons profile at D.I. Khan, Pakistan. Gomal J. Med. Sci., Vol. 16.

Kiyani, K.A., S.I. Ahmad, A. Saeed, R. Rashid and K. Abrar, 2002. Evaluation of obesity with identification of comorbidities and risk status. J. Rawal Med. Coll., 6: 82-90.

Koenker, R. and G. Bassett, 1978. Regression quantiles. Econometrica, 46: 33-50.

Koenker, R. and K.F. Hallock, 2001. Quantile regression: An introduction. J. Eco. Pers., 15: 43-56.

Leung, T.Y., T.N. Leung, D.S. Sahota, O.K. Chan, L.W. Chan, T.Y. Fung and T.K. Lan, 2008. Trends in maternal obesity and associated risks of adverse pregnancy outcomes in a population of Chinese women. Br. J. Obstet. Gynaecol., 115: 1529-37.

Marion, N. and M. Jacobson, 2000. Halting the obesity epidemic: A public health policy approach. Public Health Reports, 115: 12-24.

Martins, P.S. and P.T. Pereira, 2004. Does education reduce wage inequality? Quantile regression evidence from 16 countries. Labour Economics, 11: 355-371.

Mokdad, A.H., B.A. Bowman, E.S. Ford, F. Vinicor, J.S. Marks and J.P. Koplan, 2002. The continuing epidemics of obesity and diabetes in the United States. J. Am. Med. Assoc., 286: 1195-1200.

Nanan, D.J., 2002. The obesity pandemic-implications for Pakistan. JPMA, 52: 342- 6.

Pappas, G., T. Akhtar, P.J. Gergen, W.C. Hadden and A.Q. Khan, 2001. Health status of the Pakistani population: A health profile and comparison with the United States. Am. J. Public Health, 91: 93-98.

Peytremann-Bridevaux, I., 2007. Prevalence of overweight and obesity in rural and urban settings of 10 European countries. Prev Med., 44: 442-6.

Rehman, T., Z. Rizvi, U. Siddiqui, S. Ahmed, A. Sophie, M. Siddiqui, O. Saeed, Q. Kizilbahs, A. Shaikh, A. Lakhani and A. Shakoor, 2003. Obesity in adolescents of Pakistan. JPMA, Vol. 53.

Ruhm, C.J., 2006. Current and future prevalence of obesity and severe obesity in the United States. National Bureau of Economic Research. Department of Economics, University of North Carolina.

Siervo, M., A.A. Davies, S.A. Jebb, F. Jalil, S.E. Moore and A.M. Prentice, 2007. Ethnic differences in the association between body mass index and impedance index (Ht^2/Z) in adult women and men using a leg-to-leg bioimpedance method. Euro. J. Clin. Nutr., 61: 1337-1340.

Shah, S.M., D. Nanan, M.H. Rahbar, M. Rahim and G. Nowshad, 2004. Assessing obesity and overweight in a high mountain Pakistani population. Trop. Med. Int. Health, 9: 526-32.

WHO/IASO/IOTF., 2000. The Asia-Pacific Perspective: Redefining Obesity and Its Treatment. Melbourne: Health Communications, Australia.

Zohoori, N., T.A. Mroz, B. Popkin, E. Glinskaya, M. Lokshin and D. Mancini, 1998. Monitoring the economic transition in russian federation and its implications for demographic crisis: The Russian longitudinal survey. World Dev., 26: 1977-1993.