Empirical Project on the Determinants of a Healthy Body Mass Index (BMI)

Registration Number: 202186437

November 28, 2021

This report employs the Ordinary Least Squares (OLS) regression methodology to explain the determinants of a Healthy Body Mass Index (BMI). It is based on the 2011 Health Survey for England, a cross-sectional dataset in which 10,617 individuals were surveyed about their health, personal background, and economic circumstances. This report employs various first-order and second-order statistical tests in ensuring that the model is robust and BLUE (Best Linear Unbiased Estimators). The link to the R-workpaper can be found here: `http://www.bit.ly/319XO2D`

Contents

1 Literature Review

The BMI is calculated by dividing an individual's weight in kilogram by the square of the individual's height in meters (CDC, 2021). This index was developed to measure weight categories that may result (be associated) with health problems. Health professionals use this to check if an individual's weight is healthy (NHS, 2018).

Several authors have undertaken studies to understand the determinants of the BMI of individuals across various countries, age groups, and social strata. Lahti-Koski et al. (2000), Chhabra and Chhabra (2007), Vashishtha and Fowler (2012), Karaoglan and Tansel (2018), Mashinya et al. (2018), Nonterah et al. (2018), Puciato and Rozpara (2020) analysed data from different countries.

Chhabra and Chhabra (2007) conducted their studies in India by considering household income, sex, and place of residence. Their studies were focused on non-smoking adults and found that a problem of overweight, underweight, and obesity exists for non-smoking adults in Delhi, India. They concluded that females, economically better off, and urban residents were more likely to be overweight or obese. Simialrly, Lahti-Koski et al. (2000) analysed BMI in Finland by considering age, occupation, and education as determinants. They analysed random samples from four crossectional data sets in Finland between 1982 and 1997. They found that education was a strong determinant of obesity, especially in women, while BMI trended upwards for men outside of the labour force.

Vashishtha and Fowler (2012) extended their analysis by including race. They found that Hispanics and Blacks tended to have higher BMIs than their Asian counterparts. Their research was focused on adolescents in the United States of America. Puciato and Rozpara (2020) conducted an extensive analysis of demographic and socio-economic determinants of the working-age population in Poland. They included variables such as marital status, occupation status, income, savings, number of people in a household and employed a stepwise regression model on data collected through questionnaires and interviews.

2 Dataset & Methodology

Our data set is the Health Survey for England 2011 Dataset (University Of Manchester and Research, 2013). The dataset contains 10,617 observations with 58 variables. Our dependent or target variable is bmival. This variable represents the Body Mass Index of individuals surveyed. Table 1 below shows descriptive statistics for the BMI variable. It is interesting to note that both the mean and median are equal, and the difference between the first and third quartile is not very large when considering the range of the data. Also, there are 2,241 rows without values for BM. The regression model will automatically omit these rows.

We will employ the Ordinary Least Square regression model for our analysis. The Ordinary Least Squares (OLS) technique is intuitively appealing and mathematically much more straightforward. It has BLUE (Best Linear Unbiased Estimate) properties and shows the value of standard errors, making the technique suitable for precision tests for statistical significance. We will create dummy variables for some factors of interest, such as sex and ethnic origin.

Given our literature review on BMI's socio-economic & demographic determinants, we have itemised our explanatory variables in Table 2 in the appendices.

3 Model Specification

3.1 Functional Form of the Model

$$BodyMassIndex(BMI) = f(Sex, Age, Income, HouseholdSize,\\ EthnicOrigin, BloodPressure, AlcoholConsumption, FruitConsumption) \quad (1)$$

3.2 The Deterministic/ Mathematical Form of the Model

$$BMI_i = \beta_0 + \beta_1 Male_i + \beta_2 Age_i + \beta_3 Totinc_i + \beta_4 HHSize_i + \beta_5 Origin_i + \beta_6\\ Omdiaval_i + \beta_7 AC_i + \beta_8 FC_i \quad (2)$$

3.3 The Econometric Form of the Model

$$BMI_i = \beta_0 + \beta_1 Male_i + \beta_2 Age_i + \beta_3 Totinc_i + \beta_4 HHSize_i + \beta_5 Origin_i + \beta_6\\ Omdiaval_i + \beta_7 AC_i + \beta_8 FC_i + \mu_i \quad (3)$$

Where:

$\beta_0 =$ Intercept; $\beta_i =$ Variables specified in Table 2;
$BMI_i =$ Body Mass Index; $\mu_i =$ Stochastic Error Term

NB: We will take the natural logarithm of BMI, Totinc, Omdiaval, AC, and FC to aid in model specification and result interpretation.

4 Hypothesis

We set our hypothesis to understand which variables are determinants of BMI. Hence, we test the statistical significance of the coefficients of the variables in our model.

$$H_0 : \beta_i = 0 \text{(The coefficients are statistically insignificant)}\\ H_1 : \beta_i \neq 0 \text{(The coefficients are statistically significant)} \quad (4)$$

Decision Rule: Reject the null hypothesis if the p-value of the test statistic is less than 0.05(5%).

5 Empirical Findings and Discussion

Our model has an Adjusted R-squared of 13.5%. This result means that changes in the included determinants can explain about 13.5% of an individual's BMI changes. We use the adjusted R2 to interpret because we estimated a multiple regression model and must consider the higher degrees of freedom. While this figure is not exactly great, we must highlight that human behaviour in survey data exhibits high randomness that one model may not conclusively capture. Also, there were a lot of missing variables in our model, which caused the linear model function to drop entire rows to enable fitting.

To interpret our model, we conclude that, on average, men have a 2.3% higher BMI than their female counterparts. Also, a year increase in age will cause an individual's MI to increase by about 0.1%. Similarly, a unit increase in each household size and years of education will increase BMI on an average by 0.63% and 0.67%, respectively, all other factors held constant. Based on ethnic origin, the model shows that while White and African individuals have higher BMIs of 3.1% and 6.5%, respectively, Chinese individuals have lower BMIs on the average of about 12%.

For the logged variables, we find that while total income and alcohol consumption are negatively related to BMI, blood pressure and fruit consumption are positively related to BMI, although fruit consumption is insignificant. A percentage increase in each of these determinants will lead to a consequent 0.007%, 0.01% decline & 0.33% and 0.001 percent increase in BMI. It is worthy of note that since the dependent variable is logged, the intercept BMI is $e^{1.7828}$.

6 Evaluation of Regression Result Based on Statistical Criteria (First Order Conditions)

6.1 T - Test

Based on the result in Table 3, we reject the null hypothesis and conclude that all our coefficients are significant at the 5% confidence level. This means that our selected determinants are statistically significant determinants of an individual's BMI, except for the quantity of fruit consumed, which was insignificant in the model.

6.2 F – Test

The F-Test checks whether the variables in our model are jointly significant. This means that they are jointly different from zero. Our model returns an F-Test statistic of 59.22. The probability value is less than 5%. Hence we conclude that the model coefficients are jointly significant.

7 Evaluation of Regression Result Based on Econometric Criteria (Second Order Test)

7.1 Test for Heteroscedasticity

Heteroscedasticity tests the spread of the error terms in a model. The BLUE characteristic of OLS models requires that the error term be evenly distributed (Homoscedasticity). We test this assumption using the Breusch-Pagan test. Our model returns a BP test statistics of 130.87, significant at the 5% confidence level. We thus conclude that a problem of heteroscedasticity exists in our error terms dispersion.

7.2 Stability Test

The Ramsey Reset test is used to test for model stability. It helps us check whether non-linear combinations would work better for fitting the model. For the model fitted above, the RR test returned a value of 21.22. This value is significant at the 5% levels leading us to reject the null hypothesis and conclude that the model is unstable. Given that this is cross-sectional survey data, perhaps non-linear combinations of the variables can better explain BMI.

8 Conclusion

This report found that gender, age, household size, alcohol consumption, ethnic origin, and educational qualification are vital determinants of an individual's body mass index. These variables were significant at the 5% critical interval levels, leading us to establish a statistical relationship for the hypothesis. This work utilized an Ordinary Least Squares (OLS) method for analysis.

It should be noted that this model violates some of the BLUE assumptions of an OLS model. There is a heteroskedasticity and specification issue present. This is perhaps indicative of the kind of data that was used for analysis. There were also multiple null values across the variables, which were dropped when running the model. We recommend future iterations of the model to incorporate transformed variables, which can help with the specification and normality assumptions in the model.

Word Count: 1483

9 Appendices

Table 1: *Dependent Variable Descriptive Statistics*

Minimum	Median	Mean	Maximum	1st Quartile	3rd quartile	NA's
8.34	25.92	25.92	65.28	21.93	29.39	2,241

Table 2: *Explanatory variables included in the model*

Variable Name	HSE Variable Code	Variable definition
Male	Sex	The sex of the participant. Dummy variable encoded. Age Age Age of the participant according to last birthday
Totinc	Totalinc	The respondent's household's total income. HHSize HHSize The number of individuals in the respondent's household.
Origin (White, African, Chinese)	Origin	A demographic variable used to capture the origins of select participants in the survey.
Omdiaval	Omdiaval	Mean Diastolic of respondents. This variable measures blood pressure.
AC	Totalwu	A variable that captures the units of alcohol the respondent consumes weekly.
FC	Porfv	A variable that captures the quantity of fruit that the individual consumes.

Table 3: *Regression Result*

Variable	Coefficient estimate	Standard error	p-value
Intercept	1.7828	0.0774	< 2e-16 ***
MaleTRUE	0.0231	0.0053	1.42e-05 ***
Age	0.0016	0.0001	< 2e-16 ***
log1p(totinc)	-0.0067	0.0032	0.03724 *
HHSize	0.0063	0.0024	0.00979 **
WhiteTRUE	0.0314	0.0082	0.00014 ***
AfricanTRUE	0.0648	0.0247	0.00894 **
ChineseTRUE	-0.1201	0.0443	0.00675 **
log1p(omdiaval)	0.3327	0.0176	< 2e-16 ***
log1p(AC)	-0.0097	0.0021	3.64e-06 ***
log1p(FC)	0.0012	0.0045	0.78524
EQ	0.0067	0.0013	3.11e-07 ***

Significance Levels: ('***' 0.001), ('**' 0.01), ('*' 0.05)

Log1p: Represents variables that have been logged, inclusive of the dependent variable.

Multiple R-squared: 0.137, Adjusted R-squared: 0.1346, AIC = -3237.677

References

CDC (June 7, 2021). *Body Mass Index (BMI)*. Centers for Disease Control and Prevention. URL: https://www.cdc.gov/healthyweight/assessing/bmi/index.html (visited on 11/16/2021).

Chhabra, Pragti and Sunil K Chhabra (2007). “Distribution and Determinants of Body Mass Index of Non-smoking Adults in Delhi, India”. In: 25.3, p. 8.

Karaoglan, Deniz and Aysit Tansel (July 1, 2018). “Determinants of Body Mass Index in Turkey: A Quantile Regression Analysis from a Middle Income Country”. In: *Bogazici Journal* 32.2. ISSN: 13009583. DOI: 10.21773/boun.32.2.1. URL: http://www.bujournal.boun.edu.tr/_uploads/32_2_1.pdf (visited on 11/17/2021).

Lahti-Koski, M et al. (Dec. 2000). “Age, education and occupation as determinants of trends in body mass index in Finland from 1982 to 1997”. In: *International Journal of Obesity* 24.12, pp. 1669–1676. ISSN: 0307-0565, 1476-5497. DOI: 10.1038/sj.ijo.0801437. URL: http://www.nature.com/articles/0801437 (visited on 11/17/2021).

Mashinya, Felistas et al. (Nov. 16, 2018). “Determinants of body mass index by gender in the Dikgale Health and Demographic Surveillance System site, South Africa”. In: *Global Health Action* 11 (sup2), p. 1537613. ISSN: 1654-9716, 1654-9880. DOI: 10.1080/16549716.2018.1537613. URL: https://www.tandfonline.com/doi/full/10.1080/16549716.2018.1537613 (visited on 11/17/2021).

NHS (June 26, 2018). *What is the body mass index (BMI)?* nhs.uk. Section: chq. URL: https://www.nhs.uk/common-health-questions/lifestyle/what-is-the-body-mass-index-bmi/ (visited on 11/16/2021).

Nonterah, Engelbert Adamwaba et al. (Nov. 16, 2018). “Socio-demographic and behavioural determinants of body mass index among an adult population in rural Northern Ghana: the AWI-Gen study”. In: *Global Health Action* 11 (sup2), p. 1467588. ISSN: 1654-9716, 1654-9880. DOI: 10.1080/16549716.2018.1467588. URL: https://www.tandfonline.com/doi/full/10.1080/16549716.2018.1467588 (visited on 11/17/2021).

Puciato, Daniel and Michał Rozpara (Nov. 5, 2020). “Demographic and Socioeconomic Determinants of Body Mass Index in People of Working Age”. In: *International Journal of Environmental Research and Public Health* 17.21, p. 8168. ISSN: 1660-4601. DOI: 10.3390/ijerph17218168. URL: https://www.mdpi.com/1660-4601/17/21/8168 (visited on 11/17/2021).

University Of Manchester, Cathie Marsh Centre For Census and NatCen Social Research (2013). *Health Survey for England, 2011: Teaching Dataset*. Version Number: 1st Edition Type: dataset. DOI: 10.5255/UKDA-SN-7402-1. URL: https://beta.ukdataservice.ac.uk/datacatalogue/doi/?id=7402#1 (visited on 11/17/2021).

Vashishtha, Devesh and James Fowler (Apr. 2, 2012). “Determinants of Body-Mass Index in Adolescent Networks”. Senior Undergraduate Honors Thesis. San Diego: University of California.