Overweight and obesity are strong indicators for additional health problems (CDC, 2015). Overweight and obesity are complex issues with many cases including genetics, environmental factors, psychological elements, social dynamics, and economic issues. However, the greatest contributing factor is likely energy imbalance (Aronne, Nelinson, & Lillo, 2009). The energy imbalance results from excess dietary consumption of energy calories counterbalanced by not enough energy expended through metabolic processes and physical activity (Aronne et al., 2009). In part, this superfluous energy consumption results from an increased intake of energy-dense foods that are high in fat and sugar and a decreased intake of fruits and vegetables (WHO, 2015). American's as a whole do not consume the recommended number of fruits and vegetables per day and this contributes to the declining health of the population and increasing size of waistbands (USDA, 2015). The Body Mass Index (BMI) measure, which is calculated based on a person's height and weight ratio, is a common practice to classify someone as overweight or obese (Harvard T.H. Chan School of Public Health). In adults, overweight is defined as a BMI between 25.0 and 29.9 and a BMI greater than 30 indicates obesity (Harvard T.H. Chan School of Public Health). I chose to investigate a possible relationship between the number of servings of fruits and vegetables per day and a person's BMI. The number of servings of fruits and vegetables per day is the predictor variable, meaning it is the variable from which a prediction is made (Howell, 2011). The BMI is the criterion variable, meaning it is the variable to be predicted (Howell, 2011).

My research questions are as follows:

1.) Is there a relationship between the number of servings of fruits and vegetables a person eats per day and his or her BMI? 2.) Does the number of servings of fruits and vegetables a person eats per day predict his or her BMI?

Data collection on the variables fruit and vegetable intake per day and BMI were gathered from an IRB approved existing data set. I surveyed first and second year medical students at Ohio University Heritage College of Osteopathic Medicine through an online Qualtrics survey. For this particular research report, I selected the first 20 female responders. I used a question on fruit and vegetable intake per day and used the height and weight data to calculate BMI. BMI was calculated using the equation: weight (lbs)/ [height (in)]^2 x 703 (CDC). The raw data and descriptive statistics are below.

F/V Intake per day	BMI
4	30.7
6	20.1
3	21.1
2	24.9
3	22.6
2	23.7
3	20.1
4	28.1
2	25.1
1	31.1
3	24
6	22.7
3	28.7
4	26.7
3	19.6
5	23.3
1	25.6
3	20.2
5	20.4
3	21.3

Descriptive Statistics

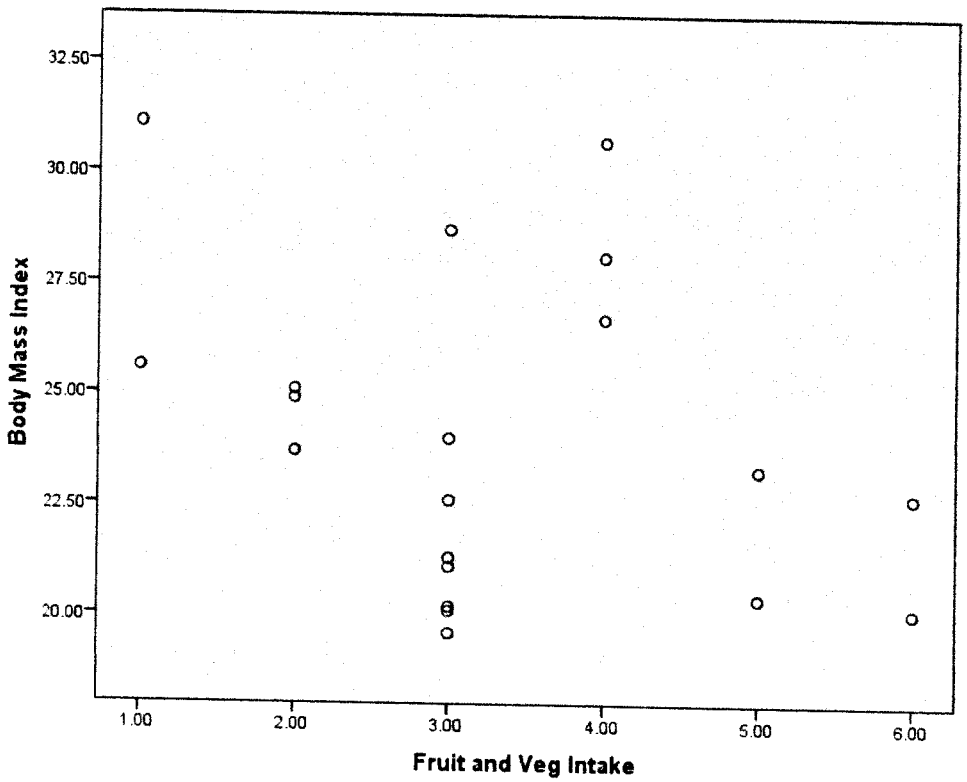
	Mean	Std. Deviation	N
Fruit and Veg Intake	3.3000	1.41793	20
Body Mass Index	24.0000	3.57756	20

Also, please find my sums table, which will be used in computations throughout the report.

Fruit/Veg Intake Per day (X)	$(F/V\ Intake)^2$ (X^2)	BMI (Y)	BMI^2 (Y^2)	F/V Intake x BMI (XY)
4	16	30.7	942.49	122.8
6	36	20.1	404.01	120.6
3	9	21.1	445.21	63.3
2	4	24.9	620.01	49.8
3	9	22.6	510.76	67.8
2	4	23.7	561.69	47.4
3	9	20.1	404.01	60.3
4	16	28.1	789.61	112.4
2	4	25.1	630.01	50.2
1	1	31.1	967.21	31.1
3	9	24	576	72
6	36	22.7	515.29	136.2
3	9	28.7	823.69	86.1
4	16	26.7	712.89	106.8
3	9	19.6	384.16	58.8
5	25	23.3	542.89	116.5
1	1	25.6	655.36	25.6
3	9	20.2	408.04	60.6
5	25	20.4	416.16	102
3	9	21.3	453.69	63.9
$\sum X = 66$	$\sum X^2 = 256$	$\sum Y = 480$	$\sum Y^2 = 11763.18$	$\sum XY = 1554.2$

Correlation

A correlation is used to show a relationship between variables (Howell, 2011). The scatterplot for the fruit and vegetable intake per day and BMI is below. The slope of the scatterplot is negative because the points on the left side of the graph are higher than the points on the right side of the graph. However, the negative slope indicates direction of relationship, not strength of the relationship.



The covariance is the statistic representing the degree to which two variables vary together (Howell, 2011).

The covariance computation is:

$$Cov_{xy} = \frac{1554.2 - \frac{(66)x(480)}{20}}{(20-1)}$$

$$Cov_{xy} = \frac{(1554.2-1584)}{19} \qquad = -1.568$$

If the covariance variable is large and positive, scores on one variable tend to be paired with high scores on the other variable (Howell, 2011). If high scores on one variable are paired with low scores on the other variable, the covariance will be large and negative. If high scores on one variable are paired equally with high and low scores on the other variable, the covariance will be close to zero (Howell, 2011). Because my covariance variable is -1.568, I can conclude that high scores on one variable correspond with low scores on the other variable.

The Pearson's correlation coefficient computation is:

$$r = \frac{1554.2 - \frac{(66x480)}{20}}{\sqrt{\left[256 - \frac{66^2}{20}\right]\left[11763.18 - \frac{480^2}{20}\right]}}$$

$$r = \frac{1554.2 - 1584}{\sqrt{[256 - 217.8][11763.18 - 11520]}}$$

$$r = \frac{-29.8}{\sqrt{[38.2][243.18]}}$$

$$r = \frac{-29.8}{\sqrt{9289.476}}$$

$$r = \frac{-29.8}{96.38} \qquad r = -0.309$$

Please find SPSS Correlation output table below.

Correlations

		Fruit and Veg Intake	Body Mass Index
Fruit and Veg Intake	Pearson Correlation	1	-.309
	Sig. (2-tailed)		.185
	Sum of Squares and Cross-products	38.200	-29.800
	Covariance	2.011	-1.568
	N	20	20
Body Mass Index	Pearson Correlation	-.309	1
	Sig. (2-tailed)	.185	
	Sum of Squares and Cross-products	-29.800	243.180
	Covariance	-1.568	12.799
	N	20	20

The negative sign of the correlation indicates the direction of the relationship: meaning as one variable is high, the other variable is low. The correlation value being -0.309 indicates not a strong linear relationship because a strong linear relationship is represented by a correlation

coefficient close to 1. The value of -0.309 does not indicate no relationship, but it does not indicate a strong relationship.

The correlation hypotheses are listed below:

H_0: $p=0$

The variables of fruit and vegetable intake per day and BMI are NOT significantly correlated in the population.

H_1: $p\neq0$

The variables of fruit and vegetable intake per day and BMI are significantly correlated in the population.

The critical value for a $\alpha = .05$, two-tailed test with 18 degrees of freedom is $r_{crit}= \pm 0.444$ (Howell, 2011). Based on the critical value, the null hypothesis is a fail to reject and I conclude that there is insufficient evidence to show that fruit and vegetable intake per day and BMI are related in this population.

The coefficient of determination measures the proportion of variance in one variable due to the relationship with the other variable (Howell, 2011). The coefficient of determination calculation is $-0.309^2 = 0.0955$. Therefore, 9.5% of the variability in BMI is predicted by the relationship with fruit and vegetable intake per day. Below, please find the SPSS output table with the coefficient of determination (R Square).

Model Summary

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.309[a]	.096	.045	3.49550

a. Predictors: (Constant), Fruit and Veg Intake

Regression

Regression is the prediction of one variable from knowledge of one or more variables (Howell, 2011). A regression equation allows for this prediction.

The slope = $-0.309\,x\,\dfrac{\sqrt{\left[11763.18-\frac{480^2}{20}\right]}{19}}{\sqrt{\frac{256-\frac{66^2}{20}}{19}}}$ Slope = $-0.309\,x\,\dfrac{\sqrt{12.799}}{\sqrt{2.010}}$

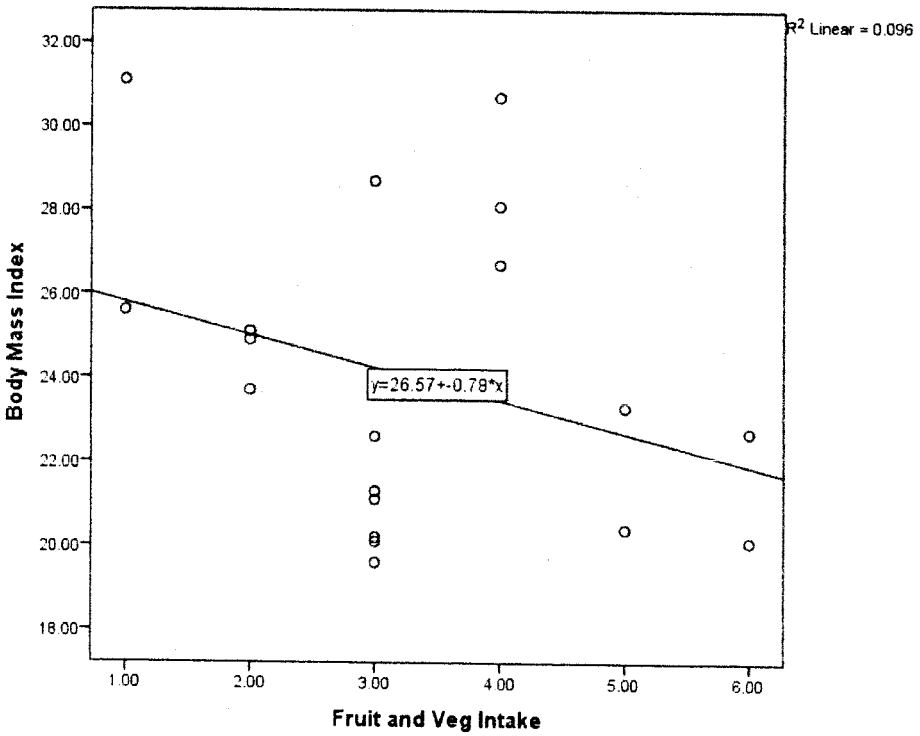
Slope = $-0.309\,x\,\dfrac{3.578}{1.418}$ Slope = -0.78

The y-intercept = $\frac{480}{20} - (-0.78)x\frac{66}{20}$

y-intercept = 24 – (-0.78) x 3.3 y-intercept = 26.57

The regression equation is: Y-hat = -0.78X + 26.57

The scatterplot with the regression equation is shown below. The coefficients table is also below.



Coefficients[a]

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta		
1	(Constant)	26.574	2.023		13.133	.000
	Fruit and Veg Intake	-.780	.566	-.309	-1.379	.185

a. Dependent Variable: Body Mass Index

Standard error of the estimate is a measure of the accuracy of predictions made using a regression line (Howell, 2011). It is measuring the distance between the regression line and the actual data points (Howell, 2011).

X	Y	Y-hat = -0.78X + 26.57	Error: Y–y-hat	Squared Error $(Y\text{-}y\text{-}hat)^2$
4	30.7	23.45	7.25	52.56
6	20.1	21.89	-1.79	3.2
3	21.1	24.23	-3.13	9.797
2	24.9	25.01	-0.11	0.0121
3	22.6	24.23	-1.63	2.66
2	23.7	25.01	-1.31	1.72
3	20.1	24.23	-4.13	17.06
4	28.1	23.45	4.65	21.62
2	25.1	25.01	0.09	0.0081
1	31.1	25.79	5.31	28.19
3	24	24.23	-0.23	0.053
6	22.7	21.89	0.81	0.656
3	28.7	24.23	4.47	19.98
4	26.7	23.45	3.25	10.56
3	19.6	24.23	-4.63	21.44
5	23.3	22.67	0.63	0.397
1	25.6	25.79	-0.19	0.036
3	20.2	24.23	-4.03	16.24
5	20.4	22.67	-2.27	5.15
3	21.3	24.23	-2.93	8.58
				SS_{error}= 219.92

Standard error of the estimate $= \sqrt{\frac{219.92}{20-2}}$ $= 3.495$

The standard distance between the actual data points and the regression line is 3.495. The SPSS output table of the model summary is below.

Variables Entered/Removed[a]

Model	Variables Entered	Variables Removed	Method
1	Fruit and Veg Intake[b]	.	Enter

a. Dependent Variable: Body Mass Index

b. All requested variables entered.

Model Summary

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.309[a]	.096	.045	3.49550

a. Predictors: (Constant), Fruit and Veg Intake

ANOVA[a]

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	23.247	1	23.247	1.903	.185[b]
	Residual	219.933	18	12.218		
	Total	243.180	19			

a. Dependent Variable: Body Mass Index

b. Predictors: (Constant), Fruit and Veg Intake

The regression equation is beneficial in the prediction of values. Subject 7 has an obtained X value of 3 and obtained Y value of 20.1. Using the regression equation, the predicted value for subject 7 is: Y-hat = -0.78(3) + 26.57 = 24.23. This value does not equal the obtained Y value. The discrepancies between the obtained Y value and the predicted Y value are accounted for with the standard error of the estimate. Additionally, the predicted Y values are just that, a prediction. They are to provide a guideline for what may occur. The obtained value is not far off

of the predicted value, but it could just be one difference in the data point, or the regression line is not completely accurate for this population.

Despite providing insufficient evidence to show a significant relationship between fruit and vegetable intake per day and BMI, it is important. Overweight and obesity are serious indicators of additional health concerns and inadequate fruit and vegetable intake contributes to this problem.