Name

Professor

Course

Date

Statistics Ratio: Body Weight and BMI

Linear Regression

When I began researching for this project, I knew I wanted data concerning body proportions. My original thought was to compare the relationship between height and weight, but that topic changed after finding a dataset through the CDC's National Health Interview Survey from 2007. This dataset included the parameters I was looking for, for almost 5000 people (the population being all people in the United States). However, as we were only required to have at least 30 data points, I limited my selection down to 100 (my sample size) which can be seen in Table 1 on the next page. I also chose to switch my topic to analyzing body weight and BMI which had the strongest correlation among the groups I compared. There were a few odd data points for those people who did not want to report a specific measure (those points were removed to try and make the analysis more accurate) (National Health Interview Survey, 2007).

In this case, my independent variable was Weight while my dependent variable was BMI. I initially expected to find a very strong relationship between the two considering that BMI is actually derived from weight. These values were all voluntarily self-reported which may affect the accuracy of the analysis. In the future these values would be more accurate if they were to be recorded and reported at a Doctor's office to confirm their accuracy from a third impartial source (CDC, 2018).

Table 1: Raw Data for 100 people's weight and BMI					
Weight (lbs)	BMI	Weight (lbs	BMI	Weight (lbs)	BMI
260	33.36	135	21.79	140	21.92
185	26.54	125	18.45	180	26.55
170	32.13	145	23.43	220	35.53
175	26.62	170	23.72	122	24.61
168	27.13	205	28.61	100	19.55
172	27.2	115	20.39	135	24.67
170	24.39	175	26.62	190	27.27
147	24.47	150	18.74	115	21.76
148	25.38	180	30.86	135	21.79
140	23.3	120	21.93	200	36.56
170	27.45	145	24.14	125	19.57
130	28.31	192	32.94	163	32.89
205	38.76	160	31.26	210	30.15
185	34.97	130	23.78	155	25.03
140	23.3	140	26.47	130	25.4
155	23.57	170	30.12	220	33.46
130	22.32	240	32.55		
160	27.46	220	29.83		
130	21	110	19.49		
165	25.82	175	27.41		
290	39.31	175	29.13		
125	23.63	158	27.12		
175	29.13	180	34.01		
105	19.84	118	19.05		
142	22.93	193	32.1		
160	27.46	189	28.73		
215	29.15	130	22.32		
160	27.46	165	23.01		
130	22.32	120	19.96		
205	28.61	140	24.81		
185	31.73	145	24.89		
225	30.52	200	27.11		
209	29.99	130	21.65		
235	37.95	174	24.27		
140	17.97	202	27.38		
225	27.41	170	29.16		
180	27.36	150	24.95		
138	22.29	190	25.77		
200	32.29	200	29.52		
235	34.69	134	21.65		
170	32.13	135	23.15		
215	36.88	148	23.89		

Table 2 below displays all of the relevant data for the linear regression run on the previous data. It was found that the correlation coefficient was

$$r = 0.8230,$$

which explains how well the data points move in unison, which for this value is pretty well. The coefficient of determination was

$$r^2 = 0.6773,$$

which explains that about 67.73% of the variation in the y (BMI) variable is explained by the x (weight) variable.

Table 2: SUMMARY OUTPUT

Regression Statistics	
Multiple R	0.82298
R Square	0.677295
Adjusted R Square	0.674002
Standard Error	2.80433
Observations	100

ANOVA

	df	*SS*	*MS*	*F*	*Significance F*
Regression	1	1617.548	1617.548	205.6833	8.3E-26
Residual	98	770.698	7.864265		
Total	99	2388.246			

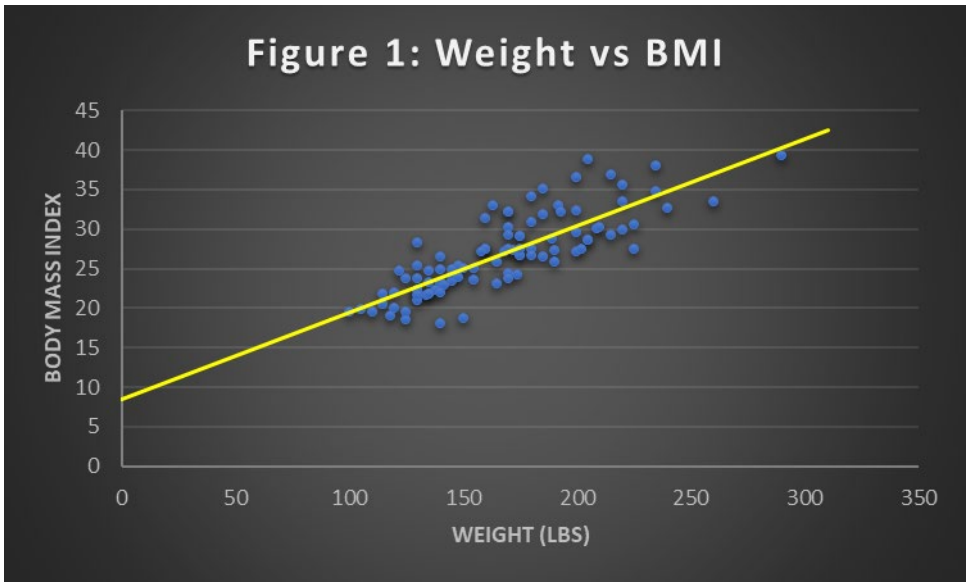
	Coefficients	*Standard Error*	*t Stat*	*P-value*
Intercept	8.43324	1.309231	6.441368	4.43E-09
X Variable 1	0.110108	0.007678	14.34166	8.3E-26

I found the regression equation to be

$$y = 8.4332 + 0.1101x.$$

The scatterplot of this data as well as the regression line can be seen in Figure 1 on the following page. As this equation is only accurate over the domain of our data [100, 290], I plugged in the

weight $x = 150$ to my regression equation on the next page to see how well it could predict BMI.



$$y = 8.4332 + 0.1101(150)$$

$$y = 24.9482.$$

Compared to the previous table, this value is pretty accurate, and even matches one of the preexisting values almost exactly. As this data is based on real world statistics, we can easily infer that this data, while now out of date, could easily be used to predict average BMI's resulting from a person's weight. However, as weight is only part of predicting BMI, height as well as musculature must be considered as well for future BMI predictions.

Confidence Interval

For calculating my confidence interval and descriptive statistics, I chose to look at Body Mass Index. There is no real reason I chose my dependent variable over my independent variable, other than it was had a smaller range than weight. I found the mean of this data to be 26.774, the standard deviation to be 4.9116, and found my 95% confidence interval to be

$(25.81, 27.74)$ which is the range of values that we can be 95% sure contain the population mean. All of these values can be seen in Table 3 below.

Table 3: Descriptive Statistics (BMI)	
Mean	26.774
Standard Deviation	4.911588029
Median	26.585
Mode	22.32
95% CI	(25.81, 27.74)

Hypothesis Test

Dividing my BMI dataset of 100 individuals into male and female groups, I hypothesize that the mean BMI of men is less than the mean BMI of women (Kuczmarski et al. 2000). So, my hypotheses are set up as follows:

$$H_0: \bar{x}_m \geq \bar{x}_f$$

$$H_1: \bar{x}_m < \bar{x}_f.$$

In this case I performed a one tail t-test with $\alpha = 0.05$. The test results are shown in Table 4 below.

Table 4: BMI T-test		
	Male	*Female*
Mean	27.373171	26.357627
Variance	17.030247	29.001522
Observations	41	59
Pooled Variance	24.115287	
Hypothesized Mean Difference	0	
df	98	
t Stat	1.0171155	
P(T<=t) one-tail	0.1558015	
t Critical one-tail	1.6605512	
P(T<=t) two-tail	0.311603	
t Critical two-tail	1.9844675	

So, I found my p-value to be 0.1558. As $0.1558 > 0.05$ ($p > \alpha$), we fail to reject our null hypothesis and conclude that there is not enough evidence to support the claim that men have lower BMI's than women. As we have failed to reject our null hypothesis, there is the risk of a Type 2 error where we have failed to reject the null when the alternative is actually true.

Works Cited

National Health Interview Survey. 2007.

Web. http://people.ucsc.edu/~cdobkin/NHIS%202007%20data.csv

CDC "National Health Interview Survey." *National Center for Health Statistics*. Centers for Disease Control and Prevention. 11 December 2018 Web. https://www.cdc.gov/nchs/nhis/index.htm Accessed on 16 December 2018.

600

Kuczmarski, R. J., Ogden, C. L., Grummer-Strawn, L. M., Flegal, K. M., Guo, S.S. Wei, R., Mei, Z., Curtin, L. R., Roche, A. F., and Johnson, C. L., "CDC Growth Charts: Unites States." *Advance Data.* 314 (2000). pp. 28. *Vital and Health Statistics of the Centers for Disease Control and Prevention/National Center for Health Statistics.* Web. Accessed 16 December 2018.