

## Final Exam – Take Home

### First Year College Student Nutrition Study

The data for your project comes from a nutritional study conducted at Youngstown State University during 1997-1998. Forty four subjects completed the study in which body measurements and nutrition data was collected at the beginning of the Fall semester and then again in the Spring semester. A portion of that data appears in the data file **nutri4a.xls** which is available at <http://csuohio.edu/holcombj/mth147/finalexam.htm>

Variable guide:

<b>studnum</b>	An identifying number to keep track of subjects
<b>gender</b>	0=male, 1=female
<b>residenc</b>	0=on-campus, 1=off-campus
<b>athlete</b>	0=non-athlete, 1=athlete
<b>totcals</b>	Total calories in the Spring semester
<b>cal3000s</b>	0=no, 1=yes for over 3000 calories in Spring
<b>cas</b>	Calcium (mg) in Spring semester
<b>ca1000s</b>	0=no, 1=yes for over 1000 mg of calcium in Spring

The main purpose of the study was to examine weight and nutrition characteristics in the college first year population. One the variables examined was the total calories consumed by subjects. The variable **totcals** gives the total number of calories. The variable **cal3000s** indicates whether the subject ingested more than 3000 calories. The USDA generally recommends calorie consumption of 2000 per day.

Another important variable is calcium. The USDA recommends that an adult obtain 1000 mg or more of calcium per day. Diet records were kept and the amount of calcium ingested is recorded. The variable **ca1000s** indicates whether a student ingested sufficient calcium.

For the following tests of Hypothesis, be sure to state the hypotheses, the test statistic, the P-value or the P-value estimate, and your conclusion. Assume that the subjects are a random sample of first year students

1. Is there sufficient evidence to conclude that the mean amount of calories (**totcalf**) is greater than 2500 calories?
2. Is there sufficient evidence to conclude that over 30% of first year students consume over 3000 calories? (Use variable **cal3000s**)?
3. Determine if **athlete** and **cal3000s** (consuming over 3000 calories) are independent.

Use the **totcals** (total calories) to predict **cas** (calcium) to answer the following:

1. Show the scatterplot.
2. Give the model for the regression line.
3. Give the  $r^2$  value, interpret its strength, and interpret its meaning.
4. Use the model to predict the **cas** when the **totcals** is 2500.
5. Are there any influential observation? Why or why not?
6. Include the residual plot. Are there any outliers? If so, how many?
7. Is **totcals** a good predictor for **cas**? Why or why not? Can you think of a way it could be improved?

Write a paragraph that describes your conclusions. Also, perform some kind of a test of hypothesis that I have not proposed. This could be a test involving a mean, a proportion, independence, or it could involve a regression analysis. Clearly state your null, alternative, test statistic, P-value, and conclusion.

---

*totcals*

---

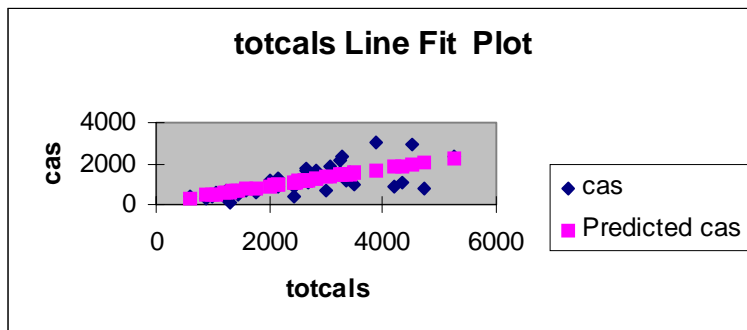
Mean 2567.664  
 Standard Error 166.0405  
 Median 2563.95  
 Mode #N/A  
 Standard Deviation 1101.388  
 Sample Variance 1213056  
 Kurtosis -0.21981  
 Skewness 0.355074  
 Range 4689.9  
 Minimum 584.5  
 Maximum 5274.4  
 Sum 112977.2  
 Count 44

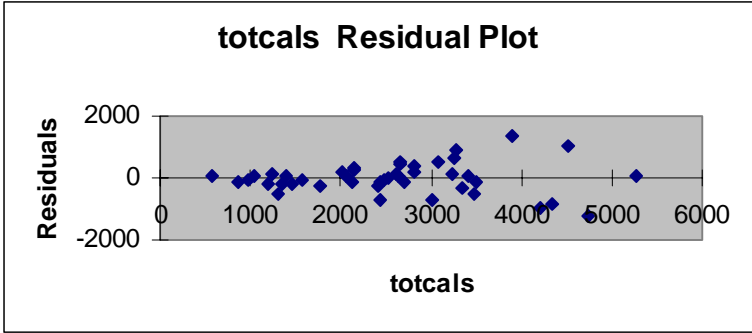
z 0.407513  
 pvalue 0.341816

Sum of count	cal3000s		
athlete	0	1	Grand Total
0	19	12	31
1	10	3	13
Grand Total	29	15	44

phat 0.34090909  
 z 0.59215653  
 pvalue 0.27687284

Chi-Square 0.996  
 df 1  
 p-value 0.318






---

*Regression Statistics*

Multiple R	0.683314
R Square	0.466918
Adjusted R Square	0.454226
Standard Error	494.8434
Observations	44

---



---

	<i>Coefficients</i>	<i>Standard Error</i>	<i>t Stat</i>	<i>P-value</i>
Intercept	86.32921	191.0899	0.451773	0.653757
totcals	0.415567	0.068516	6.065242	3.2E-07

---