Paul Dinda

Regression Analysis

Summer 2011

Student Project

The Buzz on Beer: Predicting Alcohol Content by Regression Analysis

Introduction:

In recent years, beer companies have been competing by developing low calorie beer in order to appease a more health conscious customer base. Alcoholic drinks may be consumed for several reasons; but in most cases, it is for the effect that the drug creates for the imbiber. Alcohol, a depressant, has many effects, but all of which depend on the amount of alcohol consumed. Beer contains many nutrients including: carbohydrates, protein, calcium, iron, magnesium, potassium, and niacin. This study will use regression analysis to determine if a model can be developed to predict the grams of alcohol present in a beer based on carbohydrate and calorie content.

Data:

The data was obtained on the following website http://www.beer100.com. 159 varieties of beer were studied. The serving sizes were established at 12 fluid ounces to ensure consistency in the results. The percent alcohol of the beer was converted to grams by using the known density of ethanol, 0.789 g/cm^3. All calculations were performed using Microsoft Excel Data Analysis Add-on. Every sheet was labeled per model.

Response Variable:	Y - Grams of alcohol in beer
Explanatory Variables:	X_1 - Number of calories in beer
	X_2 - Grams of carbohydrates in beer
General Model:	$Y = ß_0 + ß_1X_1 + ß_2X_2$

Model 1:

This model is a regression of the explanatory variable, number of calories, against the response variable, grams of alcohol. This will determine if the number of calories in a beer is a good indicator of alcohol content.

The single parameter equation for this model is:

$$Y = 3.4182 + 0.06944X_1$$

Regression Statistics	
Multiple R	0.805187569
R Square	0.648327021
Adjusted R Square	0.646087066
Standard Error	1.704520029
Observations	159

	Coefficients	*Standard Error*	*t Stat*	*P-value*
Intercept	3.418247327	0.598597	5.710436	5.48848E-08
Calories	0.069443239	0.004082	17.01286	1.85286E-37

The R square value in this model is somewhat undesirable at 0.6483. This indicates the model may not prove to be a great fit.

Model 2:

This model is a regression of the explanatory variable, grams of carbohydrates, against the response variable, grams of alcohol. This will determine if the amount of carbohydrates in a beer is a good indicator of alcohol content.

The single parameter equation for this model is:

$$Y = 10.1628 + 0.2767X_1$$

Regression Statistics	
Multiple R	0.432581
R Square	0.187127
Adjusted R Square	0.181949
Standard Error	2.591457
Observations	159

	Coefficients	*Standard Error*	*t Stat*	*P-value*
Intercept	10.16281	0.566891	17.92727	8.01E-40
Carbohydrates	0.276722	0.04603	6.011824	1.24E-08

This model has an extremely low R square value of 0.1871 and should be rejected.

Model 3:

This model is a regression of both explanatory variables against the response variable.

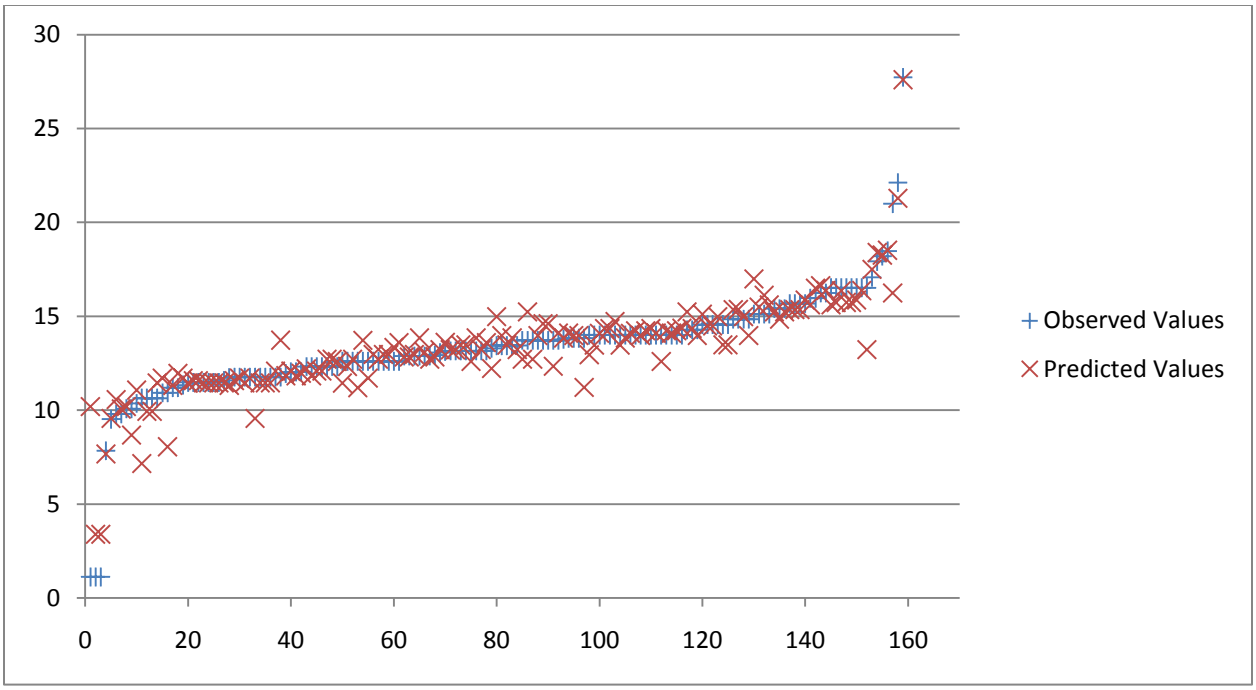
The 2 parameter equation for this model is:

$$Y = 1.1096 + 0.1260X_1 - 0.5030X_2$$

Regression Statistics	
Multiple R	0.914522631
R Square	0.836351643
Adjusted R Square	0.834253587
Standard Error	1.16647641
Observations	159

	Coefficients	*Standard Error*	*t Stat*	*P-value*
Intercept	1.109621503	0.444460516	2.496558104	0.013579393
Calories	0.126017095	0.005065536	24.87734744	3.59858E-56
Carbohydrates	-0.503015685	0.037572292	-13.38794244	1.05816E-27

The R square value, 0.8364, in this model shows that is a pretty good predictor. The P-values are quite low, which is desirable. The absolute values of the t Stat's indicate both variables are contributing significantly to the model. Plotting the observed and predicted values yields the follow graph:



Conclusion:

The best model for predicting alcohol content of a beer is model 3. It has the largest R square value, the lowest P-values, and both t Stats indicate the variables contribute greatly. Model 1 and Model 2 show that calories or carbohydrates alone do not prove as effective at determining a beer's alcohol content as the combination of both explanatory variables. Graphically, one can see that the model fits quite well except at extremely low alcohol concentrations.