Matt Doolin

Amit Singh

Kyle Thomas

Fei Xu

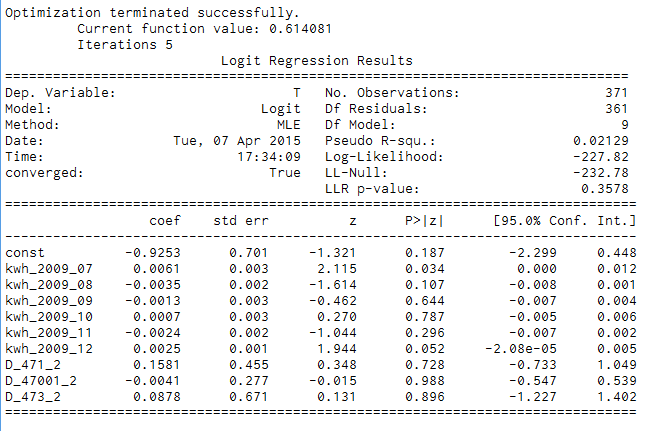
Group Assignment 3

**Way too long! You could have just said, “only A1 had significant values – but this is likely due to sampling”.**

**-1 point for lack of brevity.**

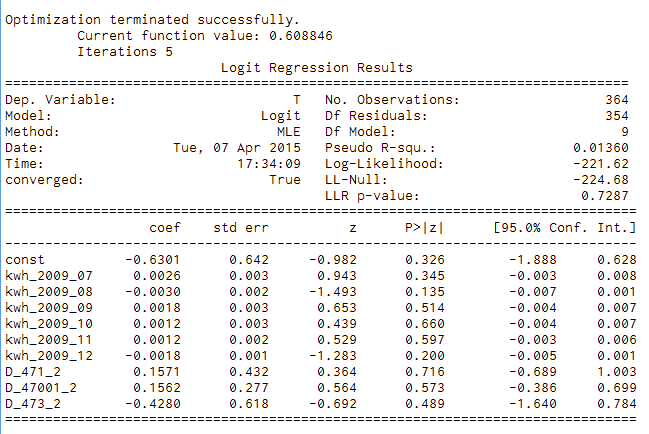
**Part I**

**LOGIT where Treatment is Tariff = A, Stimulus = 1**



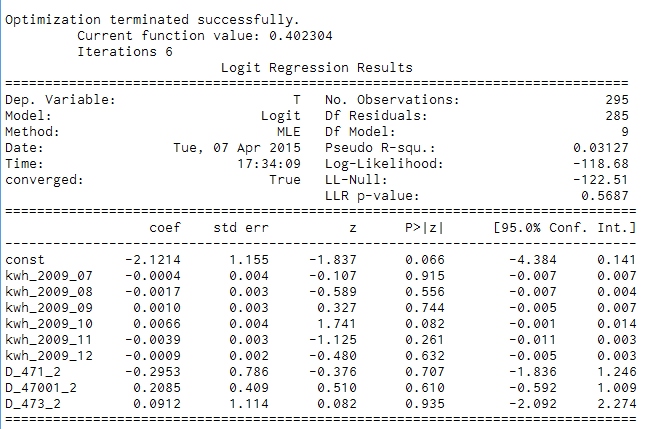
In this model, none of the coefficients are significant because their p-values are all greater than 0.05. This indicates that there is good balance between both the treatment and control groups. However, the predictor variable kwh\_2009\_07 has significant p-value of 0.034. Since this is a subsample of the entire dataset, the value for kwh\_2009\_07 could change when the entire data set is included. Additionally, given our small coefficient of 0.0061, the probability of the control group (Dependent variable = 1) having higher consumption in the July 2009 are very small.

**LOGIT where Treatment is Tariff = A, Stimulus = 3**



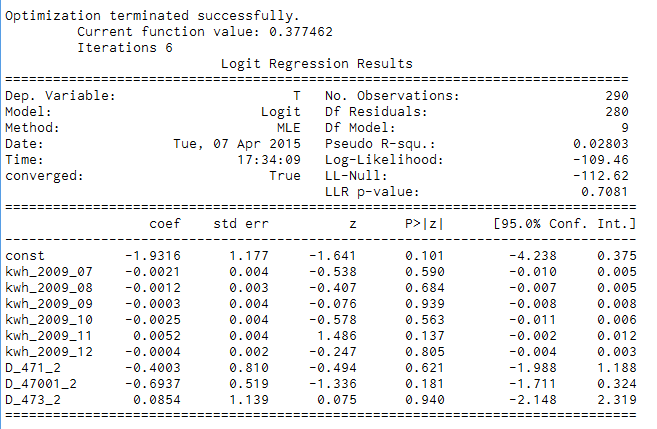
In this model, none of the survey variable coefficients are significant because their p-values are all greater than 0.05. This indicates that there is good balance between both the treatment and control groups. Additionally, none of the other predictor variables have significant p-values.

**LOGIT where Treatment is Tariff = B, Stimulus = 1**



In this model, none of the survey variable coefficients are significant because their p-values are all greater than 0.05. This indicates that there is good balance between both the treatment and control groups. Additionally, none of the predictor variables have significant p-values, although kwh\_2009\_10 has a p-value of 0.082 which is almost close to the threshold value of 0.05.

**LOGIT where Treatment is Tariff = B, Stimulus = 3**



In this model, none of the survey variable coefficients are significant because their p-values are all greater than 0.05. This indicates that there is good balance between both the treatment and control groups. Additionally, none of the predictor variables have significant p-values.

**\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_**

1. **Differences between the models:**

None of the models have predictor variables that are statistically significant. This implies that there is good balance between the treatment and control groups for all the survey questions that we chose.

**II.** **Briefly explain the benefit(s) and potential problem(s) of using ALL the available survey variables in a logit regression. Please be concise.**

Benefits: Avoid omitting key explanatory variables. Elimination of the key explanatory variables can seriously damage the explanatory power of the model and lead to biased estimates of regression coefficients.

Problems: Too many explanatory variables might lead to model overfitting, which results in variances of the estimated parameters that are larger than those for simpler models. In addition, a regression model with numerous explanatory variables may be difficult to maintain. If you include too many explanatory variables, you are bound to get one that will show an imbalance. The crucial part is determining which explanatory variables are necessary to include to properly test for balance.

Finally, the presence of many highly intercorrelated explanatory variables may substantially increase the sampling variation of the regression coefficients, detract from the model's descriptive abilities, and increase the problem of roundoff errors.

**III. Briefly explain when and why (if at all) it would be sufficient to use only a subset of the available survey questions. Please be concise.**

As we mentioned above, if you include all the survey questions, you are bound to find a question that shows imbalance. Additionally, a survey question may have a statistically significant p-value, but not provide meaning insight. Therefore, there is no reason to include the variable in the model. The key is to think about the question you are investigating, and determine which of the survey questions are most relevant to that research question.