**Ragdolls Team Project 3**

**Part I**

The results of the logit regressions are below. The questions included are gender (200), employment status (310), and internet access (405). Out of all four regressions, there are only two (non-constant) variables that are significant at the 5% level (they are the July and December kwh variables in the A1 tariff/stimulus regression).

Since each regression contains 13 (non-constant) variables, there are 4\*13=52 variables, out of which we find about 4% (2 out of 52) significant at the 5% level. That is about what we should expect if the data is completely random. This presents strong evidence of balance, since the baseline variables have essentially no predictive power for treatment group assignment.

**Part II**

The benefit of including more questions in the regression is that it “holds more things constant”. A variable might have predictive power on its own, but not when controlling for another variable.

The problem with using all available survey variables is that they often contain the same information, particularly when the survey asked multiple questions that were very similar (e.g., “do you have internet?” and “do you have broadband?”).

In certain cases, this can make it impossible to compute the model (since their parameters are not separately identified). In other cases, it can produce misleading results, like producing absurdly an absurdly large positive coefficient on one variable alongside an absurdly large negative coefficient on the other.

**Part III**

One should consider using only the subset of variables that actually matter. Adding irrelevant variables risks overfitting the model and increases the likelihood that some variables might be correlated with the error, resulting in bias.

-----------------------------------------------------------------

LOGIT where Treatment is Tariff = A, Stimulus = 1

-----------------------------------------------------------------

-----------------------------------------------------------------

Perfect Separation produced by ['D\_310\_7']. Removed.

-----------------------------------------------------------------

Optimization terminated successfully.

Current function value: 0.603307

Iterations 6

Logit Regression Results

==============================================================================

Dep. Variable: T No. Observations: 367

Model: Logit Df Residuals: 353

Method: MLE Df Model: 13

Date: Sat, 04 Apr 2015 Pseudo R-squ.: 0.04242

Time: 10:19:22 Log-Likelihood: -221.41

converged: True LL-Null: -231.22

LLR p-value: 0.1052

===============================================================================

coef std err z P>|z| [95.0% Conf. Int.]

-------------------------------------------------------------------------------

const -0.7875 0.356 -2.212 0.027 -1.485 -0.090

kwh\_2009\_07 0.0064 0.003 2.234 0.025 0.001 0.012

kwh\_2009\_08 -0.0035 0.002 -1.595 0.111 -0.008 0.001

kwh\_2009\_09 -0.0016 0.003 -0.582 0.560 -0.007 0.004

kwh\_2009\_10 0.0005 0.003 0.169 0.866 -0.005 0.006

kwh\_2009\_11 -0.0026 0.002 -1.094 0.274 -0.007 0.002

kwh\_2009\_12 0.0028 0.001 2.129 0.033 0.000 0.005

D\_200\_2 0.4508 0.236 1.910 0.056 -0.012 0.913

D\_310\_2 -0.0913 0.491 -0.186 0.853 -1.055 0.872

D\_310\_3 0.2861 0.467 0.612 0.540 -0.629 1.201

D\_310\_4 -1.4776 1.087 -1.360 0.174 -3.608 0.652

D\_310\_5 -0.1740 0.658 -0.265 0.791 -1.463 1.115

D\_310\_6 -0.4008 0.273 -1.467 0.143 -0.936 0.135

D\_405\_2 -0.0734 0.280 -0.262 0.793 -0.622 0.476

===============================================================================

-----------------------------------------------------------------

LOGIT where Treatment is Tariff = A, Stimulus = 3

-----------------------------------------------------------------

-----------------------------------------------------------------

-----------------------------------------------------------------

Optimization terminated successfully.

Current function value: 0.597312

Iterations 5

Logit Regression Results

==============================================================================

Dep. Variable: T No. Observations: 364

Model: Logit Df Residuals: 349

Method: MLE Df Model: 14

Date: Sat, 04 Apr 2015 Pseudo R-squ.: 0.03229

Time: 10:19:22 Log-Likelihood: -217.42

converged: True LL-Null: -224.68

LLR p-value: 0.4125

===============================================================================

coef std err z P>|z| [95.0% Conf. Int.]

-------------------------------------------------------------------------------

const -0.8463 0.354 -2.391 0.017 -1.540 -0.153

kwh\_2009\_07 0.0024 0.003 0.837 0.403 -0.003 0.008

kwh\_2009\_08 -0.0025 0.002 -1.203 0.229 -0.007 0.002

kwh\_2009\_09 0.0015 0.003 0.555 0.579 -0.004 0.007

kwh\_2009\_10 0.0008 0.003 0.287 0.774 -0.005 0.007

kwh\_2009\_11 0.0017 0.002 0.721 0.471 -0.003 0.006

kwh\_2009\_12 -0.0024 0.001 -1.602 0.109 -0.005 0.001

D\_200\_2 -0.0319 0.237 -0.134 0.893 -0.497 0.433

D\_310\_2 0.7245 0.455 1.592 0.111 -0.168 1.617

D\_310\_3 0.5591 0.461 1.213 0.225 -0.344 1.462

D\_310\_4 0.9673 0.547 1.769 0.077 -0.105 2.039

D\_310\_5 -0.1468 0.715 -0.205 0.837 -1.549 1.256

D\_310\_6 -0.2132 0.294 -0.724 0.469 -0.790 0.364

D\_310\_7 -0.5371 1.141 -0.471 0.638 -2.774 1.700

D\_405\_2 -0.1220 0.290 -0.421 0.674 -0.690 0.446

===============================================================================

-----------------------------------------------------------------

LOGIT where Treatment is Tariff = B, Stimulus = 1

-----------------------------------------------------------------

-----------------------------------------------------------------

Perfect Separation produced by ['D\_310\_5']. Removed.

-----------------------------------------------------------------

Optimization terminated successfully.

Current function value: 0.406859

Iterations 6

Logit Regression Results

==============================================================================

Dep. Variable: T No. Observations: 287

Model: Logit Df Residuals: 273

Method: MLE Df Model: 13

Date: Sat, 04 Apr 2015 Pseudo R-squ.: 0.03681

Time: 10:19:22 Log-Likelihood: -116.77

converged: True LL-Null: -121.23

LLR p-value: 0.7786

===============================================================================

coef std err z P>|z| [95.0% Conf. Int.]

-------------------------------------------------------------------------------

const -1.9363 0.524 -3.694 0.000 -2.964 -0.909

kwh\_2009\_07 -0.0002 0.004 -0.055 0.956 -0.007 0.007

kwh\_2009\_08 -0.0020 0.003 -0.694 0.488 -0.008 0.004

kwh\_2009\_09 0.0014 0.003 0.462 0.644 -0.005 0.008

kwh\_2009\_10 0.0062 0.004 1.615 0.106 -0.001 0.014

kwh\_2009\_11 -0.0036 0.003 -1.051 0.293 -0.010 0.003

kwh\_2009\_12 -0.0009 0.002 -0.433 0.665 -0.005 0.003

D\_200\_2 -0.1820 0.347 -0.524 0.600 -0.862 0.498

D\_310\_2 -0.1852 0.774 -0.239 0.811 -1.702 1.331

D\_310\_3 -0.0977 0.814 -0.120 0.904 -1.692 1.497

D\_310\_4 0.3897 0.841 0.464 0.643 -1.258 2.037

D\_310\_6 -0.3283 0.417 -0.788 0.431 -1.145 0.489

D\_310\_7 0.3301 1.168 0.283 0.777 -1.959 2.619

D\_405\_2 0.2891 0.409 0.706 0.480 -0.513 1.091

===============================================================================

-----------------------------------------------------------------

LOGIT where Treatment is Tariff = B, Stimulus = 3

-----------------------------------------------------------------

-----------------------------------------------------------------

Perfect Separation produced by ['D\_310\_7']. Removed.

-----------------------------------------------------------------

Optimization terminated successfully.

Current function value: 0.369148

Iterations 7

Logit Regression Results

==============================================================================

Dep. Variable: T No. Observations: 286

Model: Logit Df Residuals: 272

Method: MLE Df Model: 13

Date: Sat, 04 Apr 2015 Pseudo R-squ.: 0.05782

Time: 10:19:22 Log-Likelihood: -105.58

converged: True LL-Null: -112.06

LLR p-value: 0.4511

===============================================================================

coef std err z P>|z| [95.0% Conf. Int.]

-------------------------------------------------------------------------------

const -1.9572 0.541 -3.618 0.000 -3.018 -0.897

kwh\_2009\_07 -0.0016 0.004 -0.408 0.683 -0.009 0.006

kwh\_2009\_08 -0.0015 0.003 -0.488 0.626 -0.007 0.004

kwh\_2009\_09 -0.0016 0.004 -0.379 0.704 -0.010 0.007

kwh\_2009\_10 -0.0013 0.005 -0.278 0.781 -0.010 0.008

kwh\_2009\_11 0.0050 0.004 1.347 0.178 -0.002 0.012

kwh\_2009\_12 -0.0004 0.002 -0.209 0.834 -0.004 0.003

D\_200\_2 0.1910 0.370 0.516 0.606 -0.535 0.917

D\_310\_2 -0.8613 1.101 -0.783 0.434 -3.019 1.296

D\_310\_3 1.1366 0.583 1.950 0.051 -0.006 2.279

D\_310\_4 1.0134 0.743 1.364 0.172 -0.442 2.469

D\_310\_5 -0.2107 1.126 -0.187 0.852 -2.418 1.996

D\_310\_6 -0.2295 0.456 -0.504 0.614 -1.122 0.663

D\_405\_2 -0.3489 0.453 -0.770 0.441 -1.237 0.539

===============================================================================