Group: PUMAS

Section 2

Question 6

1. The results of the four logit models were, for the most part, insignificant. All coefficients except two (on the July and August consumption variables in the A1 model) were insignificant. The two coefficients that were significant were very small (on the order of 10^-3). This shows that consumption does not influence the placement into a treatment or control group. There is no evidence that the data is imbalanced.
2. The benefit of including all of the survey variables in a logit regression is that you can explore and determine all possible ways the data could be unbalanced. However, increasing the number of regressors increases the chances of encountering problems such as multicollinearity, or correlation between regressors, and endogeneity, or correlation between regressors and the error term. In addition, if the number of regressors is large relative to the number of observations, then you could run into the problem of overfitting. Each of these problems would make the results less valid.
3. If you were aware of any of the problems discussed in (ii), it would be necessary to use only a subset of the available survey data. For example, if you were aware that there were issues of measurement error, missing responses, or correlation between survey questions, it would be important to eliminate specific questions. Also, depending on the number of observations and the number of survey questions, it may be necessary to eliminate a subset of the survey questions.