**Report**

The used the data obtained from Juice Inc. market research to estimate the equation for juice consumption, and, therefore, help the company in understanding the market before implementation of the new product.

Starting with the descriptive statistics and graph (histogram) we can guess that most of those buying ju ice have income above 500, and those who don`t – below 1000, as well as that about 70% of those who drink juice exercise, and only 27% of those who do not, whereas there is no visible relation between consumption of juice and number of Facebook friends. Thereby we created dummies for ‘poor’ people that earn less than 500 and ‘rich’ – more than 1000 with a basis of ‘average’ people, and those dummies were interacted with log (income) to bring them to our model. At that time income was really replaced by log (income) in order to correct for inflation.

We tried tree methods: LPM (OLS), probit and logit models. In the case of the LPM there is no heteroskedasticity, while it is present in probit and logit models, but as we have seen the results do not change hugely. According to the Appendix the estimates in all three models are consistent. They have the similar coefficient signs, and same variables show statistical significance in each model. In the LPM model, pseudo R-squared is just the common R-squared from the OLS. For Logit & Probit pseudo R-squared is 0.2, thus explicative variables do not give good prediction. The number of correctly predicted is about 72%, and a lot of zeros were predicted correct as well as ones (the number of correctly predicted with only constant is about 50%).

Then, running likelihood ratio test speaks in favor of choosing the restricted model (i.e. the one without “friends” variable). Anyway the table below shows that coefficients and significance are very similar.

As for interpretation the results of the model we can state that (on average, cet. par.) exercisers should buy juice with about 20% higher probability than those who do not exercise and it is very significant. All of the models have showed the coefficient for friends to be insignificant. The impact of income is also significant and according to the different models 10% increase in income leads to about 1-2.4% increase in probability to buy/drink juice. But in the cases of Probit & Logit there is no such an average person who earns average income of less than 500, average income of more than 1000 and just average income simultaneously. Thus, we paid attention to this interpretation and suppose that in case of, for instance, Probit for average person from the middle class 10% increase in income leads to 1.84% increase in probability to buy juice, while this probability 0.17% lower for average ‘poor’ person and 0.25% higher for average ‘rich’ person.

Eventually, the more a person earns and exercises, the more he or she is likely to be buying juice.

Judging by regression outputs, it is good advice to target well-off people, and at the same time emphasize the role of healthy lifestyle in marketing campaign. Collaboration with some large fitness centers can also have success.

**Appendix**

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **LPM, Logit and Probit Estimates of Juice Consumption**  **(p-values in parenthesis, negative values in red)** | | | | | |  |
|  |
| **Independent variables** | **LPM** | **Logit** | | **Probit** | |  |
| **not-restr** | **restrict** | **not-restr** | **restrict** |  |
| Linc | 0.103 | 0.184 | 0.184 | 0.241 | 0.241 | Mr. Average  for Logit &  Probit  (special  interpretation  for *il500* &  *im1000*) |
| (0.000) | (0.000) | (0.000) | (0.000) | (0.000) |
| Exercize | 0.193 | 0.196 | 0.197 | 0.197 | 0.198 |
| (0.000) | (0.000) | (0.000) | (0.000) | (0.000) |
| Friends | 0.0002 | 0.0002 | - | 0.0003 | - |
| (0.376) | (0.432) | - | (0.366) | - |
| il500 | -0.022 | -0.017 | -0.017 | -0.010 | -0.010 |
| (0.000) | (0.000) | (0.000) | (0.103) | (0.102) |
| im1000 | 0.024 | 0.025 | 0.025 | 0.023 | 0.023 |
| (0.000) | (0.000) | (0.000) | (0.000) | (0.000) |
| Correctly predicted | - | 71.54% | 71.54% | 71.54% | 71.54% |  |
|  |
| R^2 (LPM), Pseudo R^2 | 0.252 | 0.203 | 0.203 | 0.205 | 0.205 |  |
|  |