Customer Analytics

Oliver J. Rutz

**Exercise 4 – Customer Acquisition & Scoring**

This exercise focuses on the classic scoring activity (regularly carried out for customer acquisition). The firm in question is a CD club. It can rent a list containing 500 names. The list contains 6 variables: gender, hl1, hl2, hl3, hl5, hl6, (hl… are variables computed by the list owner – as it often happens in practice, we don’t know what hl… stands for or how these variables are calculated). These variables are available for rent from the list owner and are supposed to be variables that will help in scoring customers.

The CD club has sent an invitation to all 500 names on the list to join the club. The invitation offer was a deep discount on a collector’s edition of “Pink Floyd – The Wall”. We observe whether or not each of the 500 consumers accepted the invitation: the value of *y* is 1 if the person joined the club and the value is 0 otherwise.

We use the first list of 200 persons as the **TRAINING** sample, i.e., we train (i.e., estimate) the scoring model on this data. The second list of 300 is the **TESTING** sample used to test list scoring and evaluate how successful the target selection was. The 200-person list will henceforth be referred to as the Training, and the 300-person list will be referred to as the Testing.

The data are available on Canvas as TWO CSV files. *This exercise closely follows the discussion we had in class. I strongly recommend looking at the slides when you do the exercise.*

Please follow the instructions below. ***Note: Only hand-in what is marked as “Hand-in”.***

1. Predict *y* (i.e., the decision to join the club) as a function of the available scoring variables *x* (gender and all hl…) using a **LOGIT model**. Include an intercept term to account for a base response rate. **Keep all coefficients** (i.e., do not eliminate coefficients which seems to be statistically insignificant).  
   **Hand-in:** Report coefficients and p-values.
2. Based on **your logit model**, score all individuals on in the Testing sample (you can do this manually, e.g., in Excel, or adapt the R code from class). This means calculate, for all prospects in the Testing sample, the predicted response rate.

Using your model, compute (for each individual):

* 1. *Predicted Response Rate*,
  2. *Lift***.   
     Hand-in**: Results for the first 10 Names.

1. Sort the holdout-list in decreasing order of lift.   
   **Hand-in**: Nothing
2. Plot *Marginal Response Rate* vs. *Number of Prospects Targeted*  
   **Hand-in**: Your chart
3. We know that average CLV is $30 and the solicitation cost is $12. Based on the Marginal Cost Rule determine who the CD club should send invitations to.  
   **Hand-in**: Your calculations (i.e., what’s the *Cut-Off Response Rate* and *Results* (i.e, who do you target)
4. Compute the Cumulative Sum (aka running sum) for the *Predicted Response Rates* in decreasing order. Plot the curve for curve for *Number of Positive Responses* vs. *Number of Prospects Targeted*.  
   **Hand-in**: Your chart.
5. The CD club has only 40 items of the collector’s edition of “Pink Floyd’s The Wall”. Based on the Limited Supply Rule, which prospects (and how many) on the Testing list should the CD club send an invitation to?  
   **Hand-in**: Your calculations and result.
6. Compute the Cumulative Sum (aka running sum) for the *Actual Response Rate* (recall this is either 0 or 1) in decreasing order of *Predicted Response Rate*. Plot the curve for curve for number of *Actual Positive Responses* vs. *Number of Prospects Targeted*. Superimpose on this the curve obtained in step 6 above.   
   Using the chart, comment on the differences between the *Actual Response Rates* and the *Predicted Response Rates* for the prospects in the Testing Sample. What is the impact on your results in step 7?   
   **Hand-in**: Your chart and your comments