DUKE THE FUQUA SCHOOL OF BUSINESS

Evaluating a Response Model with Lifts, Gains and ROC charts

A hard reality of response modeling is that regardless of the sophistication of the models, they are much better at sorting potentially "good" customers from the "bad" ones, than in predicting the actual probability of response. Because observed responses in a typical CRM application are usually very low, most predictive models tend to be very conservative, producing low probabilities even for training cases in which a positive response is observed. That's why response models are commonly evaluated in terms of their selectivity, or their ability to sort customers according to their true response potential.

Lifts Charts

Lifts and gains are commonly used to assess the selectivity of response models. Lift shows how well a model sorts customers in terms of observed responses, relative the average, "no model" or random performance. Lifts charts will show the expected response rate among the top deciles defined by a particular response model.

A simple way to plot the lift for a model is to first categorize its predictions or score into deciles, and to compute the observed response in each decile. This can be easily done with SPSS as I show below, using the ***Lifts & Gains Example.sav*** file. This file contains the following data for 20,000 customers:

- *Recency* - months since last purchase
- *Frequen* - total number of purchases
- *Monetary* – total dollar volume purchased
- *Response* – Response to a campaign (0=No ; 1=Yes)

Suppose you want to assess the selectivity of the *Recency* score in identifying customers who are likely to respond to the campaign using a lift table and chart:

1. Use ***Transform / Categorize variables*** to categorize *Recency* into a new variable (*Nrecency*) containing its deciles
2. Use ***Analize / Descriptive statistics / Crosstabs*** to produce a tabulation of *Nrecency* (containing the decile definition) in the rows by *Response* in the columns.

		Response to campaign		
		No	Yes	Total
NTILES of RECENCY	1	1225	286	1511
	2	2574	430	3004
	3	1361	195	1556
	4	2251	272	2523
	5	2266	216	2482
	7	2252	197	2449
	8	2288	158	2446
	9	2049	77	2126
	10	1843	60	1903
Total		18109	1891	20000

Exhibit 1 – Crosstabs of Nrecency by Response

Notice that the deciles do not have the same number of customers, as one would expect. This happens because there are many customers with exactly the same recency score.

You may then copy and paste the Crosstabs results into a spreadsheet, to compute cumulative lift, as shown below:

Recency Deciles	# Customers	Cumulative # Customers	Cumulative % Customers	# Response	Cum # Response	Cum % Response	Response Rate	Cum. Response Rate	Cumulative Lift
1 (top)	1511	1511	7.6%	286	286	15.1%	18.9%	18.9%	200
2	3004	4515	22.6%	430	716	37.9%	14.3%	15.9%	168
3	1556	6071	30.4%	195	911	48.2%	12.5%	15.0%	159
4	2523	8594	43.0%	272	1183	62.6%	10.8%	13.8%	146
5	2482	11076	55.4%	216	1399	74.0%	8.7%	12.6%	134
7	2449	13525	67.6%	197	1596	84.4%	8.0%	11.8%	125
8	2446	15971	79.9%	158	1754	92.8%	6.5%	11.0%	116
9	2126	18097	90.5%	77	1831	96.8%	3.6%	10.1%	107
10 (bottom)	1903	20000	100.0%	60	1891	100.0%	3.2%	9.5%	100
	20000			1891			9.5%		

Exhibit 2 - Lift calculations

For example, below are the calculations for decile 2:

- Cumulative # customers = 1511 + 3004 = 4515
- Cumulative % customers = 4515/20000 = 22.6%
- Cum # response = 286+430 = 716
- Cum % response = 716/1891
- Response rate = 430/3004 = 14.3%
- Cum Response Rate = 716/4515 = 15.9%
- Cumulative lift = 15.9*100/9.5 = 168

Exhibit 2 shows that the overall response rate was 9.5%. From the cumulative response rate we can see that by targeting the top 2 deciles, we would expect a better response rate (15.9%), which represents a cumulative lift of 1.68 times over the average response rate. As a larger portion of the customer base is included, cumulative lift decreases, down to no lift (average response) when all customers are contacted.

A chart depicting the cumulative lift (a scatterplot of cumulative lift by the cumulative % of customers) is shown in the next exhibit.

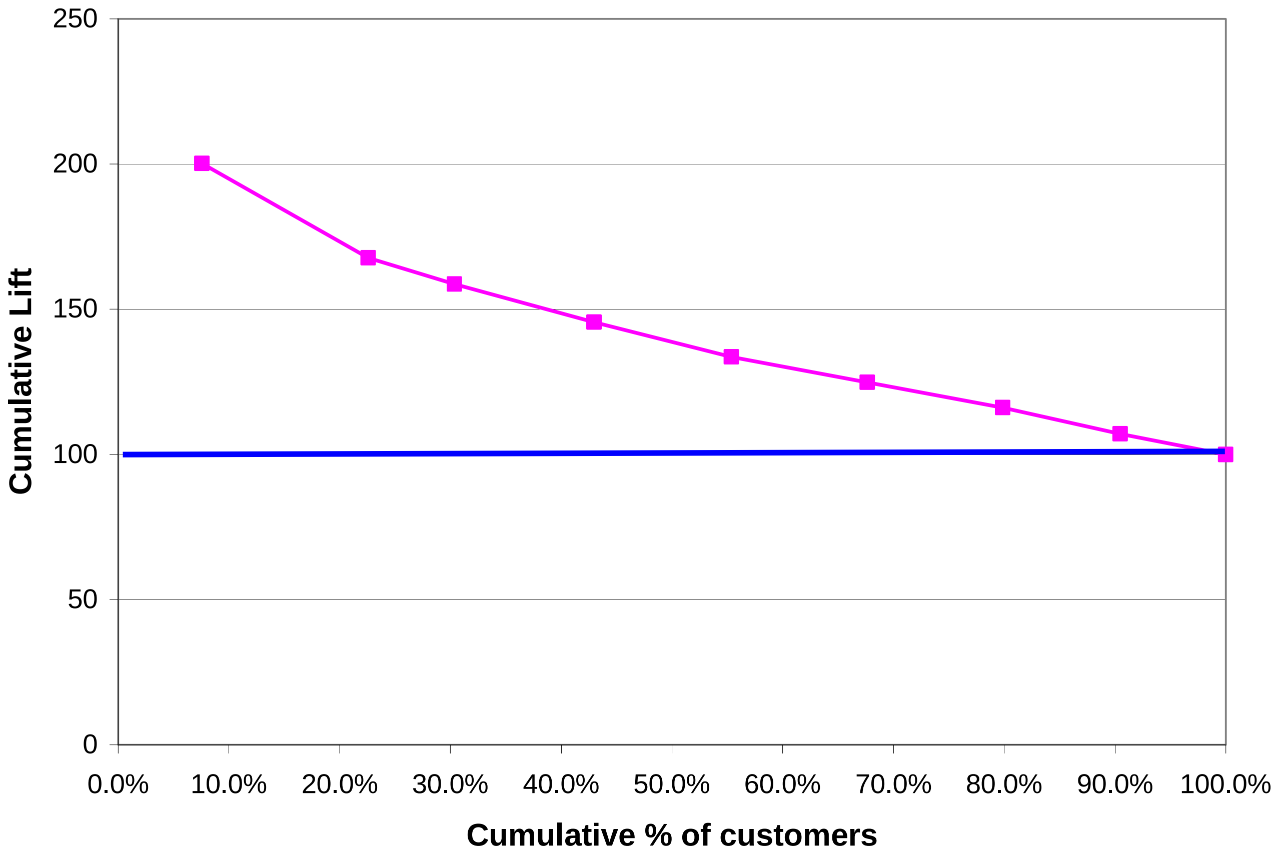

Exhibit 3 - Lift for Recency Deciles

Gains Charts

We have seen so far that *Lifts charts* show the response rate to be expected from the top deciles defined by a given response model. Another way of evaluating the response model is with a *Gains chart*. The difference is that instead of looking at response rate (% of positive responses among the selected customers) as in the lift computations, gains are defined as the proportion of respondents in the selected group, relative to all responders in the whole sample.

Starting from the same *Crosstabs* results in Exhibit 1, you can create the following table on a spreadsheet:

Recency Deciles	# Customers	Cumulative # Customers	Cumulative % Customers	# Responses	Cum # Responses	Gains	Cumulative Gains
1 (top)	1511	1511	7.6%	286	286	15.1%	15.1%
2	3004	4515	22.6%	430	716	22.7%	37.9%
3	1556	6071	30.4%	195	911	10.3%	48.2%
4	2523	8594	43.0%	272	1183	14.4%	62.6%
5	2482	11076	55.4%	216	1399	11.4%	74.0%
7	2449	13525	67.6%	197	1596	10.4%	84.4%
8	2446	15971	79.9%	158	1754	8.4%	92.8%
9	2126	18097	90.5%	77	1831	4.1%	96.8%
10 (bottom)	1903	20000	100.0%	60	1891	3.2%	100.0%
Total	20000		496.8%	1891			611.7%

Exhibit 4 - Gains and Cumulative Gains

- Gains = 430/1891 = 22.7%
- Cumulative gains = 716/1891 = 37.9%

Therefore, gains indicate the ability of the response model to identify a high proportion of the responders in the overall sample. The next exhibit shows the *Gains Chart*, also known as the “banana” chart, for obvious reason. This chart or the table in exhibit 4) shows that the top 30.4% customers ranked by *Recency* are responsible for 48.2% of all responses. The 45 degree line in Exhibit 5 indicates the selectivity obtained with the random selection of customers. The larger the distance between the gains line and this line the better the selectivity of the model is.

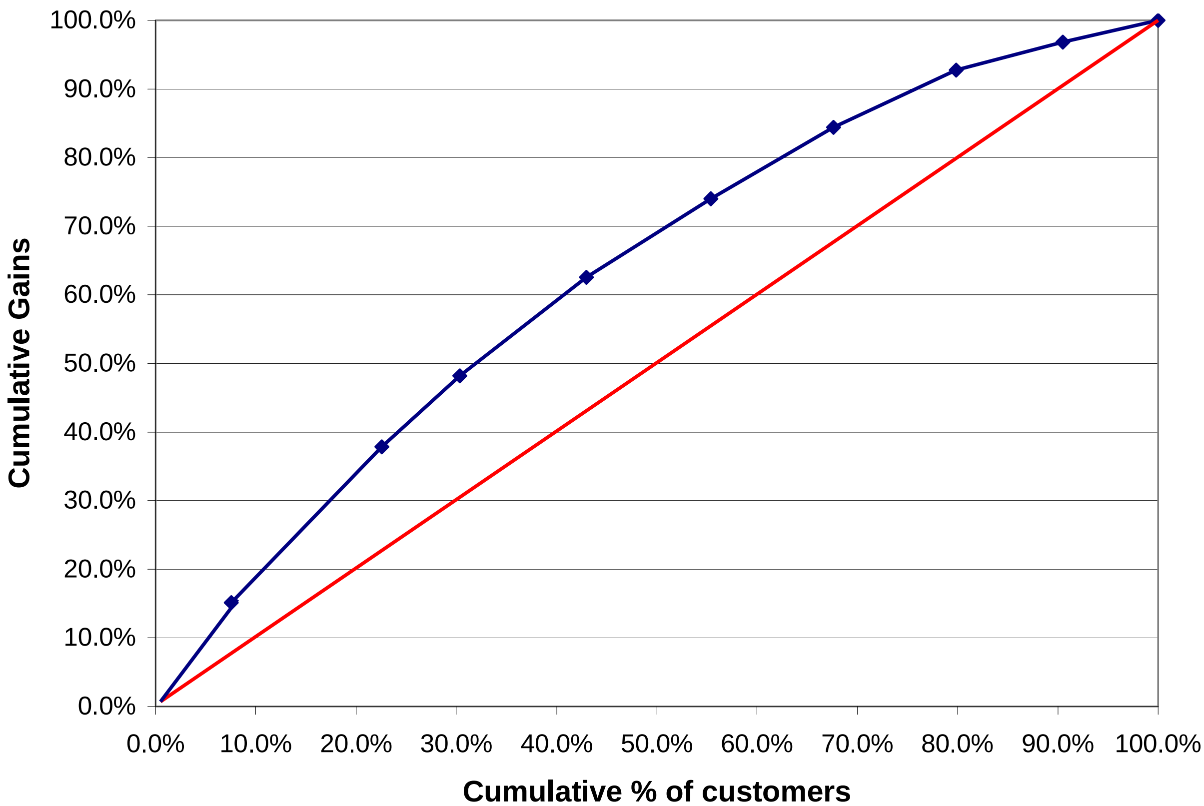

Exhibit 5 - Cumulative Gains Chart

A commonly used measure for the overall selectivity of the response model is the area between the gains line and the "no model" line. This measure of selectivity is known as the "Gini" coefficient and can be approximated by the following formula:

$$Gini = \frac{\sum_{deciles} cum\%responses - \sum_{deciles} cum\%customers}{10 - \sum_{deciles} cum\%customers}$$

In our example (Exhibit 4), the Gini coefficient is computed as:

$$Gini = \frac{6.117 - 4.968}{10 - 4.968} = 0.228$$

The Gini coefficient essentially measures the area between the Gains line and the “no model” line, relative to the total area above the “no model” line. Therefore, a Gini equal to 0 reflects the selectivity of a random selection of customers. The theoretical maximum for the Gini coefficient (when the Gains line covers the whole area above the 45-degree line) is 1.

The gains charts and tables shown above are only rough approximations for the real gains expected from the response model, because customers were grouped into only 10 “bins.” The same computations can be done on a customer-by-customer basis, replacing the 10 deciles with the actual ranking of all customers in your database. The resulting *Gains Chart* for our example data would look like Exhibit 6. The Gini coefficient in this more precise *Gains Chart* is equal to 0.272.

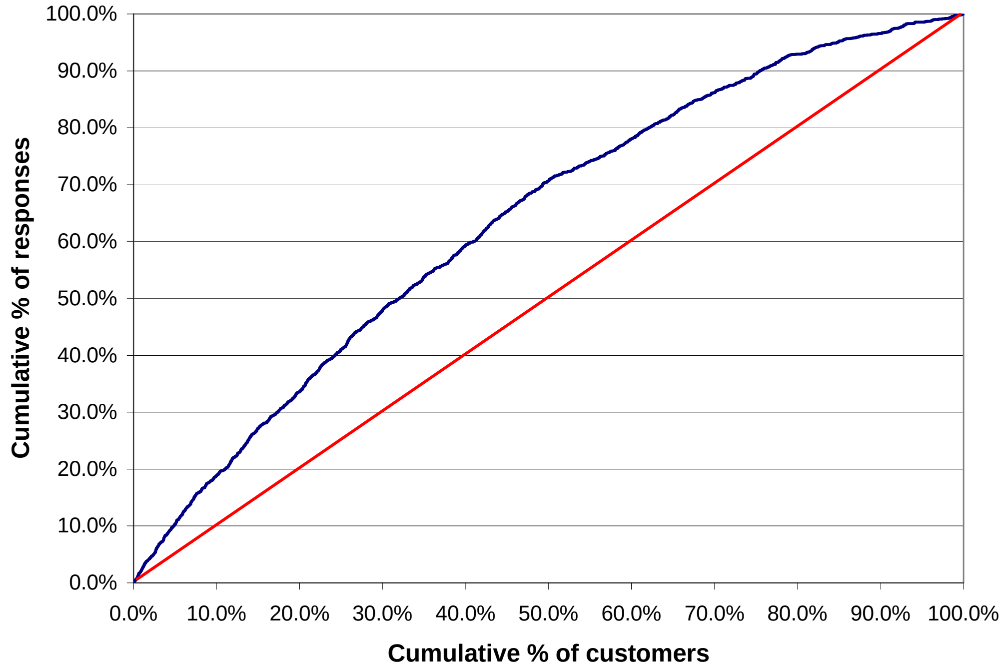

Exhibit 6- Customer-by-Customer Gains Chart

ROC Curve

The *ROC Curve,* or the “Receiver Operating Characteristics” curve was developed in signal detection theory. In CRM the ROC curve is used to assess the sensitivity of a response model in discriminate between responders and non-responders to a campaign. Suppose you have some type of score (such as the *Recency* score in our example) that is used to predict response for a large number of customers.

The first thing to be rank all customers into deciles based on their score, assuming that customers in the top deciles are more likely to respond to the campaign. The *ROC curve* works with two measures for each decile:

- *Sensitivity* : the probability that the model predicts a response when customers actually respond (of all respondents in our sample, what fraction we correctly identified as respondents?). This is nothing more than the % of all responses contained in the group of customers chosen by the model. It shows the ability of the model to "sense" responders among all responders in the sample.

- *Specificity*: probability that the model predicts no response when customers do not respond (of all non-respondents in the sample, what fraction we correctly identified as non-respondents?). This measure shows the ability of the model to rule out the non-responses among all non-respondents in the sample. Specificity is calculated as the % of all non-responses contained in the group of customers not selected by the model. In reality, ROC uses *100%- Specificity* which is also called the "false positive rate."

- A good model is one that shows high sensitivity with low false positive rates (100%-Specificity).

Confused?! Here is how *sensitivity* and *specificity* are calculated for our example:

Recency Deciles	# Customers	Cumulative # Customers	Cumulative % Customers	# Response	Cum # Response	**Sensitivity** Cum % Response	# Non-response	Cum # Non-response	**100%-Specificity** Cum % Non-response
1 (top)	1511	1511	7.6%	286	286	15.1%	1225	1225	6.8%
2	3004	4515	22.6%	430	716	37.9%	2574	3799	21.0%
3	1556	6071	30.4%	195	911	48.2%	1361	5160	28.5%
4	2523	8594	43.0%	272	1183	62.6%	2251	7411	40.9%
5	2482	11076	55.4%	216	1399	74.0%	2266	9677	53.4%
7	2449	13525	67.6%	197	1596	84.4%	2252	11929	65.9%
8	2446	15971	79.9%	158	1754	92.8%	2288	14217	78.5%
9	2126	18097	90.5%	77	1831	96.8%	2049	16266	89.8%
10 (bottom	1903	20000	100.0%	60	1891	100.0%	1843	18109	100.0%
	20000			1891					

Exhibit 7- ROC Computations

For example, for decile 2:

- Sensitivity = 716/1891 = 37.9%
- # Non-response = 3004-430 = 2574
- Cum # Non-response = 1225+2574 = 3799
- 100%-Specificity = 3799/18109 = 21.0%

The *ROC Curve* is simply a plot of *Sensitivity* against *100%-Specificity*, as shown in Exhibit 8 below:

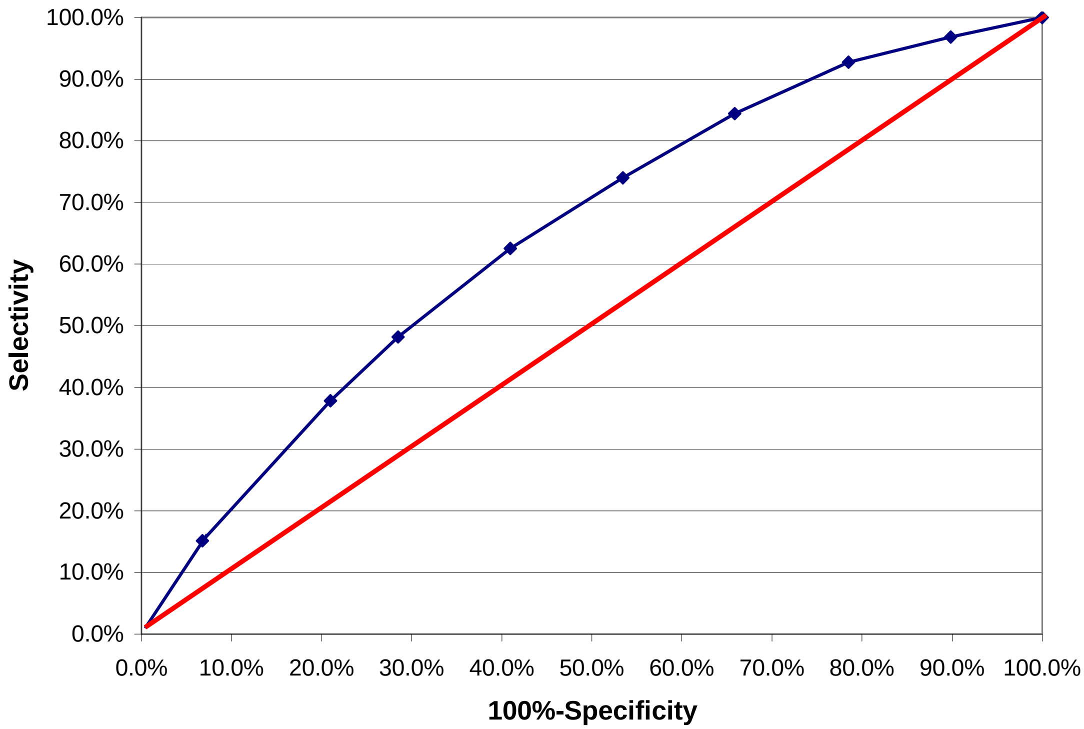

Exhibit 8 - ROC Curve

The best thing about the *ROC Curve* are that you can produce it easily using SPSS:

1. Go to *Graphs / ROC Curve*
2. Move the <u>continuous</u> predictor variable to the *Test Variable* box
3. Move the <u>binary</u> response variable to the *State Variable* box
4. Specify which *Value of state variable* will be used to count positive responses
5. Check *With diagonal reference line*
6. If your predictor is such that lower values lead to higher likelihood of response (like our *Recency* variable), click on *Options* and make the appropriate change.

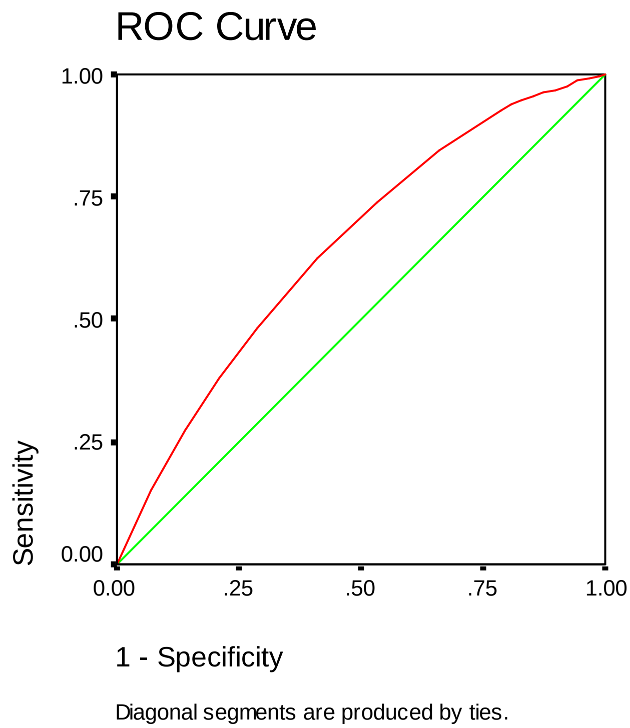

Exhibit 9 - ROC curve produced by SPSS

SPSS also produces a statistics called AUC or *Area Under Curve* which, as one would suspect, measures the total area under the ROC curve (all the way down to the x-axis). Since this *AUC* includes the triangle below the "no model" line, a value of 0.5 represents the expected performance without any model.

Area Under the Curve

Test Result Variable(s): Months since last purchase

Area
.648

The test result variable(s): Months since last purchase has at least one tie between the positive actual state group and the negative actual state group. Statistics may be biased.

Exhibit 10 - AUC statistic

Gains Charts vs. ROC Curves

Do we really need two types of charts? Don't they show the same information? Unfortunately not. ROC Curve show how accurate a particular model is in identifying positive responses with low false-positive errors. In CRM terms, we want a model that will choose the highest proportion of all responders, while at the same time avoiding making the offer to customers who will not respond.

While useful in comparing different models, ROC Curves are not very helpful in choosing the customer who should or should not be selected. They simply report error rates.

That's where Gains Charts come in handy. A Gains chart will show that if we select the x% top customers according to the model being evaluated, we will reach y% of all buyers.