

CUMULATIVE RESPONSE CURVE and ROC CURVE

There are two different (but similar curves) that can be used to assess a model's ability to rank order the data: the *Cumulative Response Curve* and the *ROC Curve*. Both curves provide similar information, but there are some differences in the information that is conveyed.

- **Cumulative Response Curves** measures how much better the model performs when compared to randomly selecting data.
- **ROC Curve** measures how well a model can differentiate between True Positives (TP) and False Positives (FP).

In the opinion of the instructor, the Cumulative Response Curve is easier to interpret. However, the ROC curve is the "Gold Standard", and most people will expect to see the ROC curve. So, for most presentations, publications, or reports it would be a good idea to include the ROC curve.

CUMULATIVE RESPONSE CURVE:

This curve compared the predictive model to randomly selection. In other words, this curve answers the question: *Is my model better than just pulling names out of a hat?* The Cumulative Response Curve is most often used in Direct Marketing applications, but it is also used in other fields.

The Cumulative Response curve is often incorrectly called a "ROC" curve. It is not a ROC curve but it does look similar which is why the confusion occurs. It is also sometimes mistakenly called a "LIFT" Chart. As with the ROC Curve, a Lift chart is something else. However, you will discover that in the real world, terminology can be flexible. In the opinion of the instructor, just figure out what the people in your organization like to call these charts and then roll with it.

Creating A Cumulative Response Curve:

1. Sort the predicted probabilities from highest to lowest.
2. Put the predicted values into "N" buckets (let's say N=10). So we put the data in 10 buckets with the highest scores in bucket 1 and the lowest scores in bucket 10.
3. Calculate the cumulative response and the cumulative difference. Here's how to do that:
 - a) Let's say that bucket 1 has 48% of all the positive responses (but if the model was random, it would only have 10% of the positive responses). So the **Cumulative Response** is 48% and the **Cumulative Difference** is 38% because 48% minus 10% is 38%.
 - b) Then we move to bucket 2. In bucket 2, we have 29% of the positive values. So that means that in the bucket 1 and bucket 2 combined we have $48\% + 29\% = 76\%$ of all the positive responses. However, if our model was random, we would have $10\% + 10\% = 20\%$ of all positive responses. So the **Cumulative Response** is 76% and the **Cumulative Difference** is 56% because 76% minus 20% random would mean that the difference is 56%.

- c) Now we move to bucket 3. In bucket 3 we might see 8% of all the positives. So we add up bucket 1+ bucket 2 + bucket 3 = 48%+29%+8%=85%. But if the model was random it would be 10%+10%+10%=30%. So the **Cumulative Response** is 85% and the **Cumulative Difference** is 55% because 85% minus 30% random would mean that the difference is 55%.
- d) We keep doing this for all 10 buckets. Incidentally, when you get to bucket 10, you will have 100% of all the positives and the theoretical value will be 100% so the Cumulative Difference in the last bucket will always be 0.

Now you have a list the Cumulative Responses and Cumulative Differences for all 10 buckets. Let's assume that the values calculated are given in Table 1. The data for this Cumulative Response Curve is provided in the Excel File included with this write up.

The maximum difference between prediction and random occurs in Bucket 2. In Bucket 2, there is a difference of 56% between using the predictive model and the randomly pulling names from a hat. So the **GAIN** of this model is **56%**. When comparing two models by using the Cumulative Response curve, the model that has the highest Gain is usually the model that is preferred. Of course common sense should be the deciding factor. Just because a model has a high Gain value, does not necessarily mean that it is a good model. Other factors need to be considered.

Note: Because the Cumulative Response Curve is sometimes incorrectly called a ROC curve, the “Gain” is also sometimes incorrectly called the “KS” value. A KS value is a similar value, but it is derived from a ROC curve. As with the ROC curve, people tend to be flexible with the terms. The instructor recommends just figuring out what people in the organization call it and conforming to the corporate lingo.

| Bucket | Cumulative Response | Random Response | Cumulative Difference |
|--------|---------------------|-----------------|-----------------------|
| 1 | 48% | 10% | 38% |
| 2 | 76% | 20% | Gain = 56% |
| 3 | 85% | 30% | 55% |
| 4 | 88% | 40% | 48% |
| 5 | 92% | 50% | 42% |
| 6 | 95% | 60% | 35% |
| 7 | 97% | 70% | 27% |
| 8 | 99% | 80% | 19% |
| 9 | 100% | 90% | 10% |
| 10 | 100% | 100% | 0% |

Table 1: Cumulative Response Data from Example

Now the Cumulative Response Curve is plotted by graphing the **Random Response Rate** on the X-Axis and the **Cumulative Response Rate** on the Y-Axis. When the data in the **Table 1** is graphed, the resulting graph is given in **Figure 1**. The diagonal “Red” line is the expected cumulative response rate if the predictions were generated randomly.

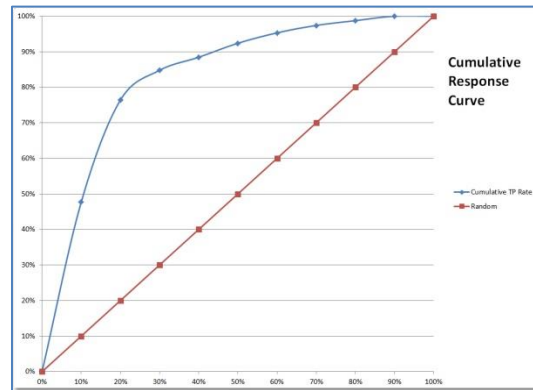


Figure 1: Graph of Cumulative Response Data

Lift Curve: In the Excel File, there is an additional graph on the Cumulative Response tab. This graph is a Lift Curve. The Lift curve simply is a ratio of the Response Rate of each bucket divided by the Random Response Rate. In this case, there were 10 buckets, so the Random Response Rate would be 100% divided by 10. This would result in a Random Response Rate of 10%. So, in the previous example, the Lift of Bucket 1 would be 4.8 (48% response divided by 10% response). The Lift of Bucket 2 would be 2.9 (29% response divided by 10%). The resulting graph is given below.

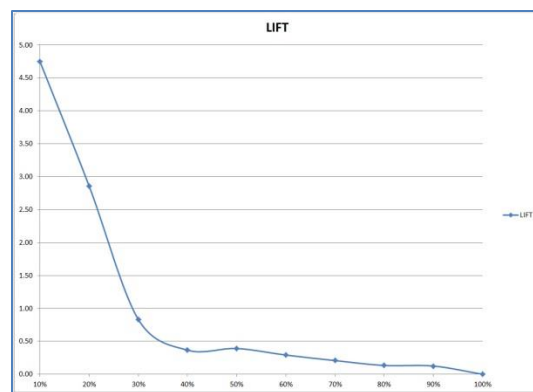


Figure 2: Lift Curve of Example Data

In Figure 2, it can be seen that after the 30% mark, the curve begins to flatten. Therefore, a direct marketer (for example) might use this chart to decide where to draw the cutoff. In this case, the graph flattens at 40% so there appears to be no advantage to using the model at this point. So the marketer might wish to score a data set and then use the top 30% of the scores for a credit card offer.

ROC CURVE:

History: the ROC curve was originally developed in the 1950's. Its original use was in the field of radar. The ROC curve was used to differentiate a radar signal from a noise signal. The graph was called the "Receiver Operator Curve" or "ROC".

The ROC curve measures how well a model can differentiate between True Positives (TP) and False Positives (FP). There are other curves similar to the ROC curve such as the previously mentioned Cumulative Response Curve, but the ROC curve is the most widely used. So it is considered to be the Gold Standard.

Creating a ROC Curve:

1. Sort the predicted probabilities from highest to lowest.
2. Put the predicted values into "N" buckets (let's say $N=10$). So we put the data in 10 buckets with the highest scores in bucket 1 and the lowest scores in bucket 10.
3. Calculate the cumulative True Positive response, the cumulative False Positive Response, and the cumulative difference. Here's how to do that:
 - a) Let's say that bucket 1 has 48% of all the positive responses and 1% of all the negative responses. So if we were to classify all the values in bucket 1 as a TRUE and every other value as a FALSE, then the **Cumulative Difference** is 47% because 48% minus 1% is 47%.
 - b) Then we move to bucket 2. In bucket 2, we have 29% of the positive values and 5% of the negative values. So that means that in the bucket 1 and bucket 2 combined we have $48\%+29\%=76\%$ of all the positive responses and $1\%+5\%=6\%$ of all the negative responses. So if we were to classify all the values in bucket 1 and bucket 2 as a TRUE and every other value as a FALSE, then the **Cumulative Difference** is 70% because 76% minus 6% is 70%.
 - c) Now we move to bucket 3. In bucket 3 we might see 8% of all the positives and 10% of all the negative values. So we add up bucket 1 + bucket 2 + bucket 3 = $48\%+29\%+8\%=85\%$ of the positive values and $1\%+5\%+10\%=16\%$ of all the negative responses. So if we were to classify all the values in bucket 1 and bucket 2 and bucket 3 as a TRUE and every other value as a FALSE, then the **Cumulative Difference** is 68% because 85% minus 16% is 68%.
 - d) We keep doing this for all 10 buckets. Incidentally, when you get to bucket 10, you will have 100% of all the positives and the theoretical value will be 100% so the Cumulative Difference in the last bucket will always be 0.

Now you have a list the Cumulative True Positives, Cumulative False Positives, and ROC Cumulative Differences for all 10 buckets. Let's assume that the values calculated are given in **Table 2**. The data for this Cumulative Response Curve is provided in the Excel File included with this write up.

The maximum difference between the cumulative true positive rate and the cumulative false positive rate is in Bucket 2. In Bucket 2, there is a difference of 70%. This value is called the **Kolmogorov-Smirnov (or “KS”) Value**. So the **KS** of this model is **70%**. When comparing two models by using the KS statistic, the model that has the highest Gain is usually the model that is preferred. Of course common sense should be the deciding factor. Just because a model has a high Gain value, does not necessarily mean that it is a good model. Other factors need to be considered.

| Bucket | Cumulative True Positives | Cumulative False Positives | Cumulative Difference |
|--------|---------------------------|----------------------------|-----------------------|
| 1 | 48% | 1% | 47% |
| 2 | 76% | 6% | KS=70% |
| 3 | 85% | 16% | 68% |
| 4 | 88% | 28% | 60% |
| 5 | 92% | 40% | 53% |
| 6 | 95% | 51% | 44% |
| 7 | 97% | 63% | 35% |
| 8 | 99% | 75% | 23% |
| 9 | 100% | 88% | 12% |
| 10 | 100% | 100% | 0% |

Table 2: ROC Curve Data from Example

Now the ROC Curve is plotted by graphing the **Cumulative False Positive Rate** on the X-Axis and the **Cumulative True Positive Rate** on the Y-Axis. When the data in the **Table 2** is graphed, the resulting graph is given in **Figure 3**. The diagonal “Red” line is the expected cumulative response rate if the predictions were generated randomly.

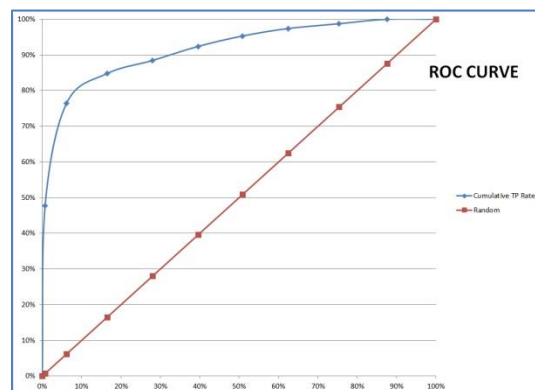


Figure 3: Graph of ROC Curve Data

Area Under ROC Curve (AUC): Another common metric that is used with the ROC curve is the Area Under the Curve or “AUC”. This is the percentage of the area that is under the curve. A perfect model would be a straight line from the origin (0%, 0%) up to the upper left corner (0%, 100%) and then another line from the upper left corner to the upper right corner (100%, 100%). A model of this type would have perfect accuracy and 100% of the area of the graph would be under the curve. A diagonal line would have

50% of the area under the curve. A visual estimate of the curve in **Figure 3** suggests that about 85% of the data is under the blue curve.

The interpretation of the AUC statistic is give in this example. Let's use the model in **Figure 3** and assume that the 85% AUC value is correct. Now, assume that a known True Positive and a known True Negative are drawn at random. Looking at the Probability score from the model for each record, there would be an 85% chance that the true positive value would be greater than the true negative value because the AUC is 85%.

Note: If the C-Statistic (not discussed in this document) is calculated, the value would be the same as the Area Under the Curve (AUC).

SIMILARITIES BETWEEN CURVES:

It is obvious by inspection that the Cumulative Response and the ROC curves are similar. The two curves have similar shapes and will generally both agree as to which model is the best. The data from the previous examples was combined on the same graph and presented in **Figure 4**. The choice of which curve to use is unimportant. The choice will likely be made based upon what others in your organization or industry are doing.

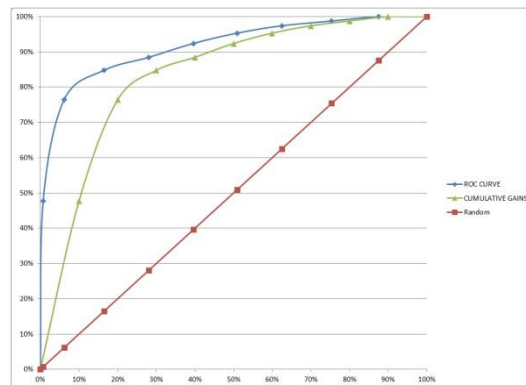


Figure 4: Combining Cumulative Response and ROC curves

INTERPRET CUMULATIVE RESPONSE CURVE

In addition to the Gain, KS, and AUC statistics, the curves can also give other information about the model. In the section below, some sample Cumulative Response and ROC curves are given and some suggested interpretations are also given.

GOOD MODEL:

A good model is shown in **Figure 5**. This model is good because there is a nice bow and all the lift is the beginning. That means that your model is good at rank ordering and when your model says that there is a high chance that a value is positive, then it is in fact positive. The extent of the “bow” in the curve will depend upon the data and the problem being solved. In some industries, a much less pronounced curve might still be considered a good model.

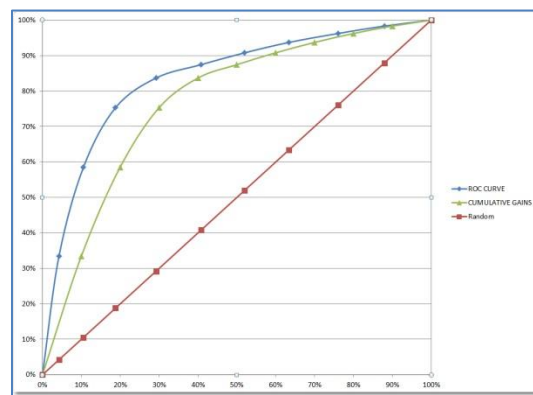


Figure 5: Good Model

TOO PERFECT:

When your model is too perfect then you need to suspect trouble. For example, examine the graph in **Figure 6**. This model is amazing. Normally, a model won't be this good. A usual cause of a "too good" model is when "future" information is accidentally used in the present. For example, if you are predicting who will file an insurance claim, you might accidentally have a value of claim amount in your model. Then your model might say, "If the person's claim amount is greater than 0 then they have a high chance of filing a claim no kidding!". So when your model is too perfect, suspect trouble! Of course, not every "perfect" model is wrong. But if the model is too good to be true, the analyst must then double and triple check their work.

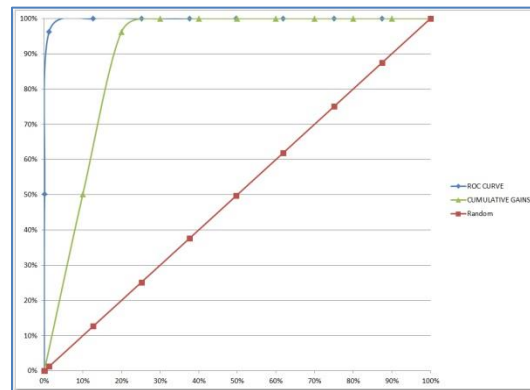


Figure 6: Too Good to be True

OVERFIT:

When you over fit your model, then your Cumulative Response and ROC curves cut through the random line as is seen in **Figure 7**. That should never happen. Your model should not ever be worse than random. Even if your Gain, KS, and AUC values are good, your model should be discarded if it cuts through the red line. Your model is no good. Count on it!

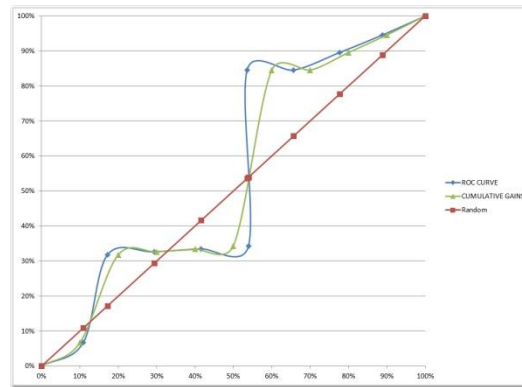


Figure 7: Overfit Model

WORTHLESS:

If your model doesn't look much different from the random diagonal line as with the graph in **Figure 8**, then it's probably a bad model. Your model should always be a lot better than pulling names out of a hat. That's why they are paying you!

Of course, maybe there just isn't anything in the data to work with and this is the best you can do. Then maybe the model given is better than nothing. However, when you get a model like the one in **Figure 8**, then you need to make certain that you have tried everything. Maybe this is a time to try some strange transformation or an exotic modeling tool like a random forest or a neural network. Regardless, if you have a graph like the one below, then better make certain that this is the best you can do.

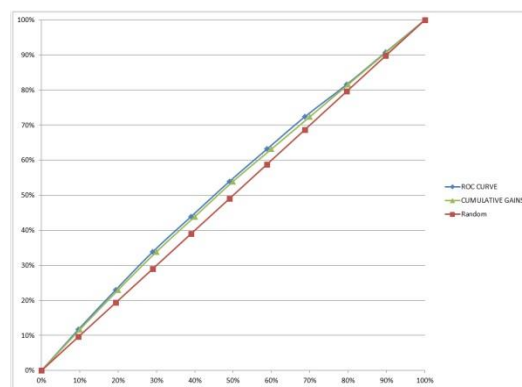


Figure 8: Too Bad