Jacob Dineen

MBC 638

Data Collection and Analysis Paper

March 25th, 2017

**Problem Definition:**

Since I am currently working in the space of digital advertising, I would like to focus my efforts on RoAS, or Return on Ad Spend. There is not necessarily a prevalent issue I’m dealing with now, but there are some kinks in our programmatic online display advertising, as far as not hitting certain key performance indicator benchmarks. This seems to be a process that is thought to be under control, but my feelings are inclined to believe that the results are not being maximized to their fullest extent. To measure performance in advertising, the key metrics that are focused on are generally clicks, click through rate, conversions, conversion rate, and all-encompassing cost calculations surrounding them- For example, conversion rate is the total conversion count divided by the total click count. The Trade Desk, a demand side platform, will be the focus of my analysis. Different sites and targeting mechanisms (proprietary data usage) are available when working with The Trade Desk as opposed to advertising through the Google Display Network, which appears to be a process that is already in control in terms of RoAS. The problem, as it currently stands, is the conversion rate measured using this platform versus the conversion rates using a platform like Google or Facebook for online display advertising. I know I have a problem because the data is telling me that out of all the consideration/conversion vendors/platforms we run on, this is the highest aggregate CPA, or cost per action, and the lowest conversion rate. With much of the data being hidden, or learned over time, as opposed to opt ins through Google or Facebook, I wouldn’t necessarily consider these to be on the same playing field, so it’s better for me to look at performance as an isolated instance, rather than comparing it too heavily across different platforms. This problem, as we’ll call it throughout this paper, needs to be fixed because our performance is based on driving these conversions, which is essentially a click out from our site to a specific dealer site- For context, our site is a representation of a regionalized groups of automotive dealers who pool their money together. We act as a portal from the Tier 2 automotive level to the Tier 3 sites/dealers, which are actual dealerships- Getting someone from our site to any of the dealers’ sites is registered as a conversion. It’s our main KPI and can be associated loosely with lead generation. If we aren’t driving success in the form of our main key performance indicators, there isn’t justifiable intelligence that leads us to continue the use of the platform. Because this is purely a test of which X variables are driving the changes in my Y variable, I will pivot back to using a variation of my initial Y, but in a continuous manner, rather than a pure count- i.e. conversion rate- and use subsidiary conversion rates by each of my categorical breakouts as a continuous measure of my inputs. Most of the data that I will be using, if not all of it, will be continuous. The great thing about my process is how easy it is to adjust once I have a statistical understanding of which inputs are truly impacting my output. If I see certain variables are stalling or derailing my overall outcome, I can simply toggle a switch to ensure that these variables don’t continue to run. When these switches, or levers, are toggled, it ensures that the spend that is no longer going into these predictors will now feed into the others. I will measure my CR (Conversion Rate: Clicks/Total Conversions) against a variety of variables- This will be my Y in this equation. A conversion is an expansion of a click- It is an action that a user takes after initially clicking on an ad and being directed to a new page. What I will look at is the relationship between my conversion rates in sub variables of device type, ad size, and fold, as well as some additional metrics that are of a continuous nature. These additional variables don’t generally have a strong relationship with my outcome, based on some high-level analysis I’ve performed in the past, but this will assign an actual statistical relevance to what I have only inferenced at eyeball level. What I hope to do through my actions and multivariate testing is have a direct impact on the overall conversion rates of the campaigns that I am currently running. I would deem any positive direction a success, but because there is a standard deviation present in the daily data, I will shoot for somewhere between a 10-20% increase in mean conversion rate, using historical data from the previous month as a benchmark. For example, I am currently generating around a 5% conversion rate day over day, meaning that my ads are driving 5 out of 100 people to engage not only with the ad, but with the designated landing page they click through to. If I could influence that final conversion rate by 10%-20%, I would see an increase to 5.5% or 6%, overall**.** Unable to attribute purchase data back to advertising efforts I must rely specifically on the ad dollars that are being saved, and that are/would be available if this DMAIC process were to render positive results. If my conversion rate increases by 10%, there is a direct impact on the cost per individual conversion. The impending impact would result in 10% more dollars to spend during any given month to pursue further leads- This equates to roughly 6,000$ per month. I think that there is difficulty in calculating the monetization of process improvement due to a grey area which falls beyond predictability – For example, if the client is thrilled with the results, we could see an increase in the budgets for this platform, which would lead to higher profitability and margins from a business perspective, but for the sake of this process, we’ll use the above value as a ROI on process improvement.

**Operational Definitions:**

RoAS: Return on Ad Spend. Like ROI.

Impression: The act of an ad being shown/Eyeballs on an ad.

Click: The act of clicking/tapping on an ad.

CTR: Measure of success (KPI)- Clicks/Impressions

Conversion: Clicking on a button/link after the initial click to the landing page.

* + Click through conversions: Conversions directly attributed to the advertisement shown.
  + Post view conversions: Conversions indirectly attributed to the advertisement shown. This is counted when a user clicks an ad, but does not convert. If the user is then exposed to a different form of advertising and converts, this registers as a postview conversion.

Spend: The amount of money spent.

Conversion Rate: Measure of success- Conversions/Clicks

Fold: Above or below- Where the ad is being shown on a page.

Size: The size of the banner/JPG being advertised. This analysis will only look at the following three sizes. 320x50 is a popular mobile size, while the others can generally be served across devices due to their larger size:

* + 320x50
  + 728x90
  + 300x250

Device type: The device medium that the ad is being absorbed on. This analysis will only look at the following three device types:

* + Mobile/Smartphone
  + Tablet
  + PC/Desktop

Win Rate: The number of impressions won/The number of impressions bid on

I think my boundaries, or some of my current limitations are correlated with the number of variables that are in play. There are so many different data points that can directly impact performance, and my hope is to be able to isolate the main drivers of performance, or to be able to weight them in a way to makes optimization more strategic, rather than just spraying and praying. The downfall is that some of this data can be tedious to collect, so I will be picking a handful of variables that I believe will be the most impactful on my Y output. The main boundary I am faced with is in relation to conversion counts. Because I don’t generate a statistically significant amount of click through conversions through this medium of advertising, I must rely on click through + post view conversions, as defined above, to perform a relevant analysis. Lastly, sometimes there just isn’t enough device information, whether it be due to issues with user cookies or IP masking, to correctly classify data under the umbrellas that I have established for this assignment. That data is instead filtered into a category that are classified as ‘unknown’. Because I can’t influence that unknown category (meaning I can’t shut off spend to an unknown lever), I will exclude it from calculations pertaining to this assignment.

I currently run all of operations that occur at the programmatic level, so the results are directly indicative of the work I’m putting in. I think this helps me better understand the process because there are no extra measures being added in for human behavior, besides my own. Not many people at my company are data savvy, so this is a task I’m putting on my shoulders for the time being, although I am more than willing to adjust tactics based on knowledgeable intelligence.

**Define**- The problem/process has been defined within the scope of my comprehension. This step is a work in progress throughout the lifetime of my analysis, but the main structure is currently in place. Operation definitions have been established, as well as the business impact of improving my process. The scope of the process has been defined, as well as the types of data that I am working with.

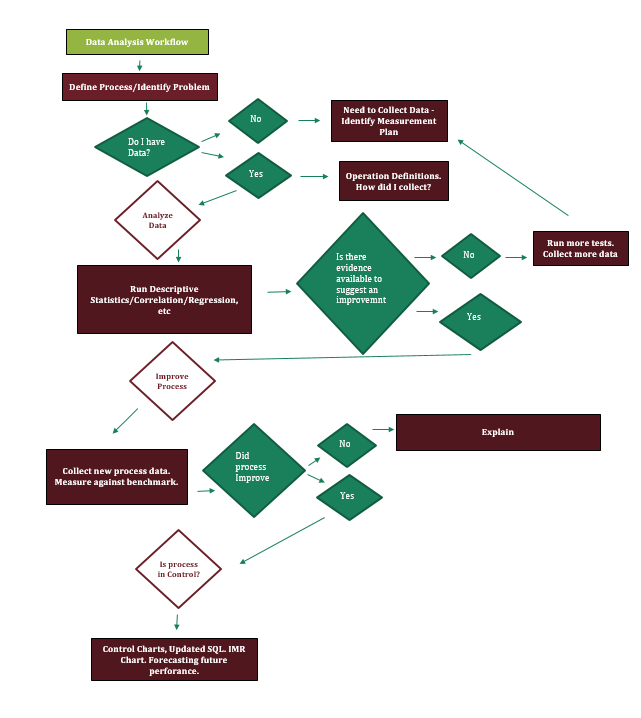
**Measure**- I am lucky to be able to use historical data through this portion of DMAIC. I will rely on the past month’s data (January) to measure performance/KPIs as I have defined above. The measurement part of my process will rely solely on data that has already occurred, and will be what I use to run my tests. The measurement data that I am drawing upon will be my ‘before’ process, and will be referred to as such throughout this project.

**Analyze**- I will run various tests over the course of a week to determine which variables are statistically significant- This step will use the data that I mentioned in the measure phase above, and will focus on the past month’s metrics. The analyze stage of my DMAIC process will include basic visualization, correlation, regression, standard deviation/variance, and process control charts, among others. These should help me to identify parts of my process that are prohibitive to my Y output, as well as helping me to understand what is noise, and what is a signal. I will need to measure and analyze the process after it goes through a change to understand the new process and my impact on it.

**Improve**- I will implement the findings of my analysis towards the live campaigns- This will essentially be turning off current predictor variables that are deemed the least valuable to my output. Then I will measure performance over the course of a two week stretch to provide a large enough sample. The improve part of my process will essentially be the February data I collect since I must set up new campaigns for the new month.

**Control**- The control stage of my process does not end when this assignment does. I must continue to make minor tweaks to sustain performance. I would like show some types of process control charts, visualizing upper and lower limits so that I may understand when my process is signaling that something is awry. The process, if all goes well, will be under control for a long enough period of time to chart. Once I can chart the new process, I can identify my control limits and be signaled when my process is breaking control.

**Process Roadmap:**



**Baseline Measurement**

Because I have historical data that I could use to analyze my process, I referred to the prior month’s data as my baseline to conduct analysis, and to benchmark against. The data from the date range January 10th through February 5th will act as my benchmark data. This will be the information that I run statistical analysis on. Once that analysis is complete, the data from February 6th to a future date will act as my new process data, or the data associated with my process after changes were implemented.

**Data Collection:**

I already have structured data to work with for this analysis, as granted to me through the reporting APIs of the platforms I am working with, so my data collection plan does not require me to collect any new data, but instead relies strictly on my ability to format/slice it in a way that makes sense for said analysis. Most of the work I will be doing is done through excel, and the same goes for the statistical analysis that I will be performing. I have a running totals spreadsheet where I will refresh and pull in data through the prior day once a week. Most of these metrics come from the same place, but some of them must be manual created via reporting features, and some of the percentages transfer over more clearly when I customize a pivot field for them. Below is a screengrab of my data collection process/methods:

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Performance Measure** | **Data Source/Location** | **How Will Data be Collected** | **Who Will Collect Data** | **When Will Data be Collected** | **What Range will Data fall within** |
| Impressions, Clicks, CTR Conversion Rates, Bids, Win Rate, Spends, Clearing Prices | TTD Reporting/API. Manual Excel Pivots/Calculations | TTD Reporting/API. Manual Excel Pivots/Calculations | JD | Feb 6th | Jan 10th- Feb 5th (Baseline Data) |
| Impressions | Tradedesk Reporting API | TTD Reporting/API. Manual Excel Pivots/Calculations | JD | Feb 13th | Feb 6th-Yesterday |
| Clicks | Tradedesk Reporting API | TTD Reporting/API. Manual Excel Pivots/Calculations | JD | Feb 13th | Feb 6th-Yesterday |
| Conversions | Tradedesk Reporting API | TTD Reporting/API. Manual Excel Pivots/Calculations | JD | Feb 13th | Feb 6th-Yesterday |
| Conversion Rates | Tradedesk Reporting API | TTD Reporting/API. Manual Excel Pivots/Calculations | JD | Feb 13th | Feb 6th-Yesterday |
| Spends by Device, Format, Fold | Tradedesk Reporting API | TTD Reporting/API. Manual Excel Pivots/Calculations | JD | Feb 13th | Feb 6th-Yesterday |
| CTR, Win Rate Calculations | Tradedesk Reporting API | TTD Reporting/API. Manual Excel Pivots/Calculations | JD | Feb 13th | Feb 6th-Yesterday |
| Impressions | Tradedesk Reporting API | TTD Reporting/API. Manual Excel Pivots/Calculations | JD | Feb 20th | Feb 13th-Yesterday |
| Clicks | Tradedesk Reporting API | TTD Reporting/API. Manual Excel Pivots/Calculations | JD | Feb 20th | Feb 13th-Yesterday |
| Conversions | Tradedesk Reporting API | TTD Reporting/API. Manual Excel Pivots/Calculations | JD | Feb 20th | Feb 13th-Yesterday |
| Conversion Rates | Tradedesk Reporting API | TTD Reporting/API. Manual Excel Pivots/Calculations | JD | Feb 20th | Feb 13th-Yesterday |
| Spends by Device, Format, Fold | Tradedesk Reporting API | TTD Reporting/API. Manual Excel Pivots/Calculations | JD | Feb 20th | Feb 13th-Yesterday |
| CTR, Win Rate Calculations | Tradedesk Reporting API | TTD Reporting/API. Manual Excel Pivots/Calculations | JD | Feb 20th | Feb 13th-Yesterday |
| Impressions | Tradedesk Reporting API | TTD Reporting/API. Manual Excel Pivots/Calculations | JD | Feb 27th | Feb 20th-Yesterday |
| Clicks | Tradedesk Reporting API | TTD Reporting/API. Manual Excel Pivots/Calculations | JD | Feb 27th | Feb 20th-Yesterday |
| Conversions | Tradedesk Reporting API | TTD Reporting/API. Manual Excel Pivots/Calculations | JD | Feb 27th | Feb 20th-Yesterday |
| Conversion Rates | Tradedesk Reporting API | TTD Reporting/API. Manual Excel Pivots/Calculations | JD | Feb 27th | Feb 20th-Yesterday |
| Spends by Device, Format, Fold | Tradedesk Reporting API | TTD Reporting/API. Manual Excel Pivots/Calculations | JD | Feb 27th | Feb 20th-Yesterday |
| CTR, Win Rate Calculations | Tradedesk Reporting API | TTD Reporting/API. Manual Excel Pivots/Calculations | JD | Feb 27th | Feb 20th-Yesterday |
| Aggregate of All Metrics | Tradedesk Reporting API | TTD Reporting/API. Manual Excel Pivots/Calculations | JD | Mar 1st | Feb 1-Feb 28 |

\*JD denotes Jacob Dineen (Who)

These are the descriptive statistics for the ‘before’ part of my process. This is essentially what I will be benchmarking against, particularly the mean seen below. The Stdev and observation counts below will also help me to know if my proposed sample is large enough to provide statistical relevance.

|  |  |
| --- | --- |
| *Conversion Rate* | |
|  |  |
| Mean | 0.054695725 |
| Standard Error | 0.004013695 |
| Median | 0.055988315 |
| Mode | #N/A |
| Standard Deviation | 0.020068477 |
| Sample Variance | 0.000402744 |
| Kurtosis | 0.427314139 |
| Skewness | 0.781830668 |
| Range | 0.077373131 |
| Minimum | 0.029200574 |
| Maximum | 0.106573705 |
| Sum | 1.367393132 |
| Count | 25 |

Mean= 5.49%

Median= 5.59%

Standard Deviation= 2%

|  |  |  |  |
| --- | --- | --- | --- |
| *Point* | *Conversion Rate* | *Rank* | *Percent* |
| 7 | 10.66% | 1 | 100.00% |
| 8 | 8.65% | 2 | 95.80% |
| 5 | 8.52% | 3 | 91.60% |
| 10 | 8.12% | 4 | 87.50% |
| 2 | 7.04% | 5 | 83.30% |
| 9 | 6.34% | 6 | 79.10% |
| 11 | 6.24% | 7 | 75.00% |
| 14 | 6.03% | 8 | 70.80% |
| 6 | 5.98% | 9 | 66.60% |
| 12 | 5.76% | 10 | 62.50% |
| 1 | 5.76% | 11 | 58.30% |
| 3 | 5.67% | 12 | 54.10% |
| 23 | 5.60% | 13 | 50.00% |
| 24 | 5.07% | 14 | 45.80% |
| 13 | 5.04% | 15 | 41.60% |
| 20 | 4.72% | 16 | 37.50% |
| 4 | 4.67% | 17 | 33.30% |
| 18 | 4.21% | 18 | 29.10% |
| 16 | 3.68% | 19 | 25.00% |
| 21 | 3.50% | 20 | 20.80% |
| 22 | 3.40% | 21 | 16.60% |
| 15 | 3.22% | 22 | 12.50% |
| 25 | 2.97% | 23 | 8.30% |
| 19 | 2.95% | 24 | 4.10% |
| 17 | 2.92% | 25 | 0.00% |

**Sample Size**

Because I am using historical data, I have the flexibility to increase my sample at will, but advertising is largely impacted my seasonality, so I want to use data from Q1 of 2017 for both my before and after process. That being said, I have 25 samples (my samples are e quivalent to days recorded in this process) to measure against.

I am dealing with continuous data – Measuring conversion rates by day.

I want to have 95% confidence in the result – I will use a z\* of 1.96

My margin of error (E) can be within 1%.

|  |  |
| --- | --- |
| **Same Size for Continuous Data** |  |
| Z\* | 1.96 |
| StDev | 2.01% |
| Margin of Error | 1% |
| Same Size Calculation | 15.47 |

This means 16 samples are needed to detect a change in the population mean. Our test data includes observations from 25 samples, meaning that we likely have enough samples available to be able to detect a change in the population mean.

If we wanted to measure our sample size with a smaller error than the 1% that was used above, we would need to increase the number of samples taken.

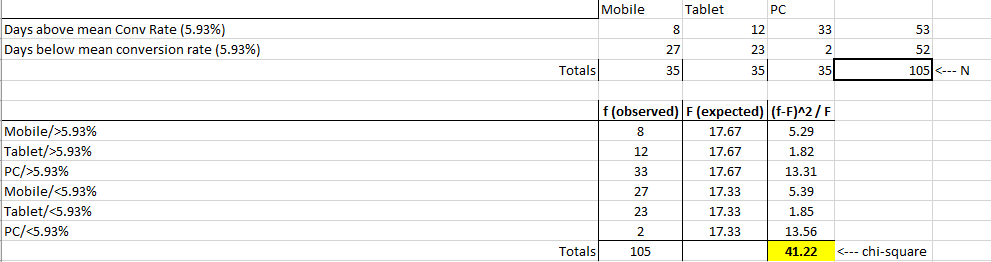
**Measurement error**

There are several measurement errors that could have occurred throughout this stage of DMAIC. Most of them are highly specific to advertising, but some of them are simply filtering errors, or including dirty data. For example, there is a possibility that video campaigns were included in the filter. These are entirely different KPIs and would distort the weight of display campaigns. There is also a possibility that some of the more manual calculations could be wrong, whether that be due to ‘fat fingers’, or using the wrong fields. Lastly, there could have been measurement error when looking at the daily conversion rates. I tried to eliminate days were activity was low so that I didn’t distort the averages too much – For example, if something registered a conversion rate 3x the average, but only spent 10% of the normal budget. This would lead to misconceptions about the trends of my data, and would ultimately distort the analysis.

**Hypothesis Testing**

Performing the Chi Square Hypothesis tests on variables that I am wanting to test will let me know if I am on the right track. I want to know if there is a relationship between my output, conversion rates, and my inputs, which are measured separately as device type and ad size.

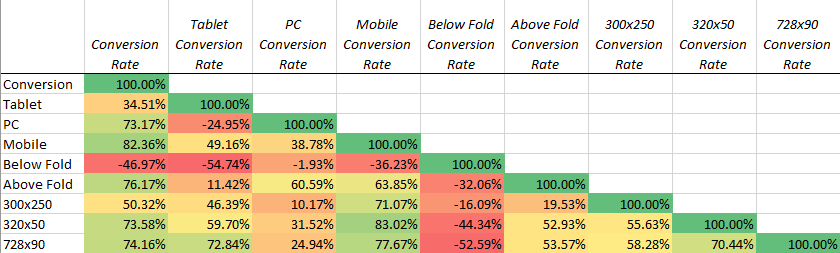
My null hypothesis would assume that these variables are independent and there are not relationships present, and my alternative would assume that there is a relationship present, but the direction is not known through the test. To measure these variables against the output, I split up conversion rates into different segment falling under the conventions of good and bad- Good being greater than the consolidated mean output, and bad below that same threshold- Note that I relied on a greater sample digging back into December 2016 for more statistical relevance.



The result was a P Value of almost zero, signifying I should drop the null hypothesis and accept the alternative, which is that these variables are related to one another.

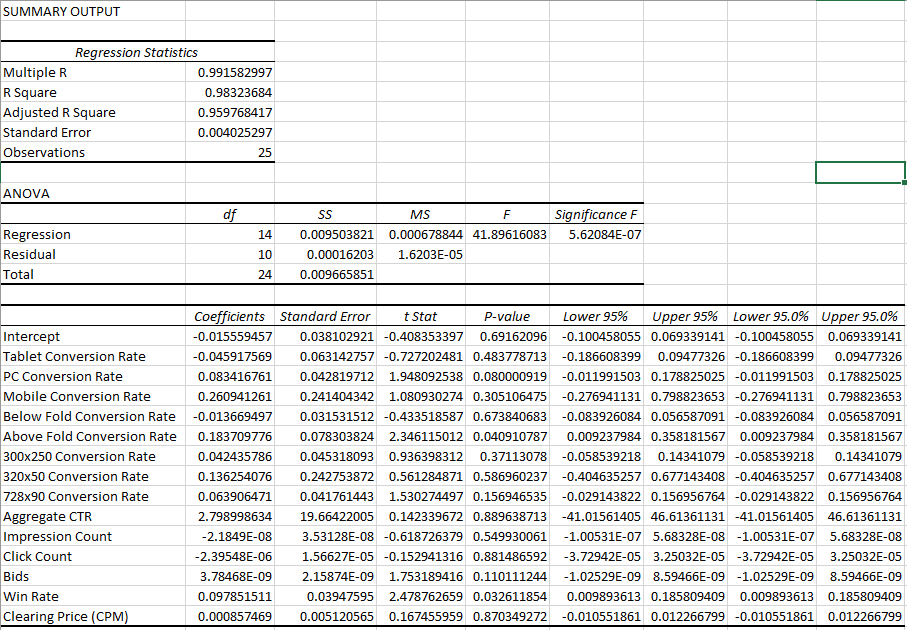
My other hypothesis test measuring ad size against conversion rates also rendered similar results, so I know that I am on the right track, or that I am at least performing a relevant analysis.

**Correlation**



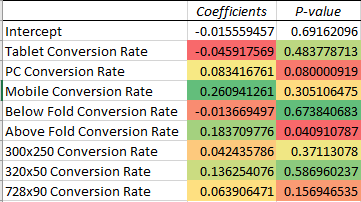
Running a correlation test only tells me that the device type = tablet, and the fold = below are negatively correlated with my output. The rest show reasonably strong relationships with variance in the aggregate conversion rate. What this also tells me is correlation amongst my variables, which showcases collinearity. But correlation does not equal causation, so this test simply provides a piece of the bigger picture.

**Regression**



Regression is a more telling indicator of statistical significance among variables and the variance in the output that corresponds to them. I ran the regression above with all my data, which is likely why my R^2 is so high. I don’t think that building a regression model will be particularly useful to me since I don’t have forecasts for predicted values to plug in to the formula, but I can draw from the pvalues above to determine some insight. Basically, I am seeing that PC is the most significant device type, while Fold = Above is the most significant fold, and 728x90 and 300x250 are the most significant ad sizes. I can’t necessarily throw variables out just because they have a pvalue greater than alpha=0.50, but I can judge which ones are more relevant than others using pvalues and coefficients in more of an eyeball approach, and draw upon my own inference to best determine a change in the process. As many of our articles have expressed, a good analysis is determined by a mixture of analytical tools and business acumen. Because I understand my process more than someone else might, I can draw upon that knowledge and combine it with the actual analytics.

The data is essentially reinforcing my beliefs prior to this experiment. Some basic scatter plots will show me that PCs, above the fold, and larger ad units result in higher conversion rates than the alternative.



On to my process. Because I have some knowledge into the actual statistical drivers of my Y, I can adjust my process. To start, I am going to limit my predictor variables as follows:

Device Type: Shift a majority of spend into PC.

Ad Format/Size: Shift a majority of spend into 300x250 and 728x90 ad sizes.

Fold: Shift a majority of spend into ‘Above the Fold’

I think that to perform a more accurate regression analysis on this data set, I would need to expand my predictor variables, probably using binary expansion, and would likely need to rely on a probit or logit analysis because my data isn’t linear. I think that introducing moderating effects would also be beneficial, because I am bound to have some form of collinearity present amongst my predictors that I would want to scrub before conducting a thorough analysis.

**Results**

*Before Process Change:*

*After Process Change- Feb 6th through Feb 26th:*

*Total Process- Jan 10th through Feb 26th:*

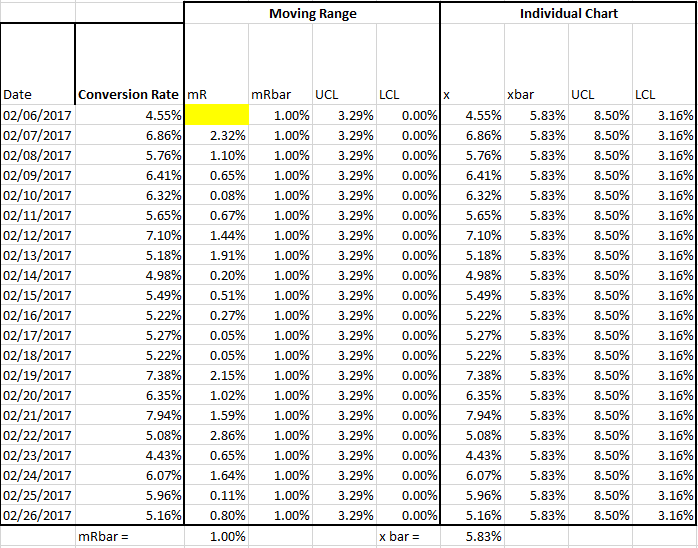
|  |  |  |  |
| --- | --- | --- | --- |
| Before | | After | |
| *Conversion Rate* |  | *Conversion Rate* |  |
|  |  |  |  |
| Mean | 0.054695725 | Mean | 0.058283935 |
| Standard Error | 0.004013695 | Standard Error | 0.002034637 |
| Median | 0.055988315 | Median | 0.056540489 |
| Mode | #N/A | Mode | #N/A |
| Standard Deviation | 0.020068477 | Standard Deviation | 0.009323877 |
| Sample Variance | 0.000402744 | Sample Variance | 8.69347E-05 |
| Kurtosis | 0.427314139 | Kurtosis | -0.135748426 |
| Skewness | 0.781830668 | Skewness | 0.659357685 |
| Range | 0.077373131 | Range | 0.035100152 |
| Minimum | 0.029200574 | Minimum | 0.044320138 |
| Maximum | 0.106573705 | Maximum | 0.07942029 |
| Sum | 1.367393132 | Sum | 1.223962643 |
| Count | 25 | Count | 21 |

\*Measures of location and dispersion seen above for before and after process change.

I will focus mostly on basic statistics to determine if my change in the process had a positive impact. As seen above, my average conversion rate, at a daily level, increased 6%. This falls slightly below my own goal, and could possibly be due to additional levers that could have been added in as predictors, such as time of day, or even things such as creative content, which is a much harder aspect to quantify. Perhaps more important than my average are the additional statistics noted above, such as min/max, range, and standard deviation. These metrics all moved in a more positive direction that indicates my process isn’t quite as wonky, and my output should be more predictable moving forward.

I learned that my process is slightly more complicated that I originally thought it would be. I thought this would be a relatively simple fix to increase my Y output by shutting of bad predictor variables, but that wasn’t necessarily the case. From what I can tell through my knowledge acquired over the duration of this course, I did have a direct impact on my process improving, but there are some things that I probably could have done better, or some additional indicators that I could have amplified my focus on. My output should likely be an extension of activity so that extra weight isn’t pushed onto days with low activity but high performance.

**Control**



The process, as outlined above using an IMR chart, is under control. I can use the Individual chart above to know when my process start to go out of control – it will signal me when my conversion rates cross the lower control limits, specifically when a day of performance data falls below 3.16% (LCL), which will tell me that my process needs to be improved or adjusted.

**SQL Before and After**

The SQL of my process is difficult to express because I am looking at a specific part of my process, and not taking a holistic view of the process. If I were looking at it as a whole, my defects might be defined as mistakes in the campaign setup process which could be prohibitive to campaign performance. That wouldn’t necessarily make sense here, so I will arbitrarily define a defect as a day of performance data which falls below one standard deviation of the mean (This average will be determined through splitting the collected data into ‘before’ and ‘after’ segments.)

**Before:**

Defect Opportunities in Process: 25

Total Possible Defects in Process: 25

Total Actual Defects: 3

Defect Per Opportunity Rate: .12

DMPO: 120,000

SQL: 2.7

Yield: 88.50%

**After:**

Defect Opportunities in Process: 21

Total Possible Defects in Process: 21

Total Actual Defects: 2

Defect Per Opportunity Rate: .09

DMPO: 90,000

SQL: 2.8

Yield: 90.30%

My Sigma Quality Level increased after the change to my process, which indicates that the process improved.

**Forecast (Moving Average and Exponential Smoothing):**

Because I have my data formatted, I could autogenerate moving averages and exponential smoothing averages to help me forecast for future performance. This would likely be more useful if I could structure it by week, or by month to help me forecast a longer range.

