**CAUSAL CONTROL CHARTS**

**Question 2**

We are trying to determine whether the 2016 Presidential Election had an effect on the percent of days that Humana stock went up.  In solving this problem, we will utilize the switch method to create a causal control chart.  We will use a dataset that contains data for 60 days from before the election (the controls) and 60 days after the election (the cases) for a total of 120 days.  A sample of which is provided in this slide.  Notice the difference between pre and post elections.  The control limits are calculated from pre-election and used to judge the changes in post-election.



We are interested in whether Human stock went up if we remove the effect of the general economy as measured by Nasdaq, and the health care sector as measured by S&P Health stock.  There are two sets of these stocks one for pre-election and another for post-election. There could be other factors known to affect the value of Humana’s stock, but for the purposes of this exercise we are assuming that these two factors are all that matters.  We refer to Nasdaq and S&P as alternative explanations for why Humana’s stock goes up or down.

Control limits for a p-chart are calculated from the average probability of the event among the controls.  To calculate the **Upper Control Limit (UCL)** and the **Lower Control Limit (LCL)**we need to measure the average probability of Humana stock going up prior to the election. First, we show how this calculation is done without adjusting for Nasdaq and S&P healthcare stock prices.

**Step 1: Break up our dataset into time periods**

First, we are going to break up our dataset of 120 records (60 pre-election and 60 post-election) into time periods.  Since we have 60 records pre-election, we will create 10 time periods consisting of 6 days the stock market was open.  Note that the days in our dataset may skip days, since the stock market is not always active daily (such as on holidays and weekends).  Additionally, the number of time periods is arbitrary and you may for the most part choose a different number of time periods.

**Step 2: Organize the strata**

For a complete breakdown of the combinations in our dataset, let’s look at all the possible combinations (or strata) for the two alternative stocks that we are considering (S&P Health, and Nasdaq).  The goal of our calculations is to make sure that the average probability of Humana going up is calculated in such a manner that before and after election will have the same trends in S&P healthcare and Nasdaq.   Please note that in our data, whether each stock went down (indicated as 0) or up (indicated as 1) is binary and therefore can *only* be one of two values (either 0 or 1).

|  |  |
| --- | --- |
| **Nasdaq** | **S&P** |
| 0 | 0 |
| 0 | 1 |
| 1  | 0 |
| 1 | 1 |

As you can see, there are four possible combinations and we want these four combinations to occur as frequently before and after election. All calculations of the adjusted control limits are done using these combinations.

**Step 3: Calculate the P-Average using the unadjusted frequencies from the pre-election (controls) data**

Let us perform our analysis without adjusting for the impacts of Nasdaq and S&P Healthcare.  We start by counting the number of days in each strata and dividing it by the total number of days in the pre-election period (60 days).  This value is called the *unadjusted frequency*, since no adjustments were made to this frequency.  We then take the total number of days in each strata when the Humana stock was up and divide it by the total number of days in the pre-election period to get the rate the Humana stock went up.  We sum these four rates (one for each strata) which gives us the **P-Average.**  The formulas in this slide show how each of these variables are calculated.

|  |  |
| --- | --- |
| **Strata** | **Using Pre-Election (Controls) Data** |
| **Strata Count when Humana Up** | **Unadjusted** **Frequency** | **Rate Humana Stock went Up** |
| **Nasdaq Down and S&P Down**(Nasdaq=0 and S&P=0) | =SUMPRODUCT(1-C3:C62,1-E3:E62,G3:G62) | =SUMPRODUCT(1-C3:C62,1-E3:E62)/COUNT(G3:G62) | =Q16/COUNT(G3:G62) |
| **Nasdaq Down and S&P Up** (Nasdaq=0 and S&P=1) | =SUMPRODUCT(1-C3:C62,E3:E62,G3:G62) | =SUMPRODUCT(1-C3:C62,E3:E62)/COUNT(G3:G62) | =R16/COUNT(G3:G62) |
| **Nasdaq Up and S&P Down** (Nasdaq=1 and S&P=0) | =SUMPRODUCT(C3:C62,1-E3:E62,G3:G62) | =SUMPRODUCT(C3:C62,1-E3:E62)/COUNT(G3:G62) | =S16/COUNT(G3:G62) |
| **Nasdaq Up and S&P Up** (Nasdaq=1 and S&P=1) | =SUMPRODUCT(C3:C62,E3:E62,G3:G62) | =SUMPRODUCT(C3:C62,E3:E62)/COUNT(G3:G62) | =T16/COUNT(G3:G62) |
| **P-Average** | =SUM(S16:S19) |

Keep in mind that for binary data, the SUMPRODUCT function calculates the number of times two or more variables co-occur.  For example, the formula in the last row calculates the number of times Nasdaq was up, S&P was up and Human went up.  All three variables must be 1 before the product of the values will be 1 and count towards the total sum. If any of the three variables is 0 then the product is 0 and it does not add to the sum.

The above calculations performed in Excel produce the following:



In the excel file, **$S$20**is the P-Average. Taking the **P-Average**, we can calculate the UCL in Excel as  **=IF($S$20+3\*SQRT($S$20\*(1-$S$20)/6)>1,1, $S$20+3\*SQRT($S$20\*(1-$S$20)/6))** and the LCL is calculated as **=IF($S$20-3\*SQRT($S$20\*(1-$S$20)/6)<0,0,$S$20-3\*SQRT(($S$20\*(1-$S$20))/6))**. The if statement adds the clause that if the UCL is above 1 then round it to 1. The next if statement adds the clause if the LCL is below 0 then round it to 0 since LCLs cannot be negative.   The **LCL** and the **UCL** value are below:



**Step 4: Calculate the observed frequency for each time period using *post-election* data**

Using the post-election data, break up the data into 10 time periods of 6 days each.  For each time period, calculate the observed frequency by counting the number of days the Humana stock was up in the time period divided by the total number of days in the time period (which in our case is 6).  Do this for each time period.  In Excel, you can do this using the SUM and COUNT functions.  You will do this for each of the 10 time periods and will get something like the following:



After charting these results in Excel, we get the following.



Note that the wide variation in stock prices has produced control limits that are wide apart.

The way to interpret this chart is as follows:  The control limits show the rate of Humana stock based on the pre-election situation. None of the post-election stock changes has values exceeding or falling below the pre-election control limits. Upon looking at the data, it appears as though the election results did not impact the Humana stock.

However as previously mentioned, our calculations may have been impacted by confounding variables or alternative explanations.  In other words, did the effects of the general economy and/or the healthcare industry affect our results?  If so, we want to remove these effects and **ONLY** consider the impact the election had on the Humana stock.

In order to remove the effects of the confounding variables, we have to make a few modifications to calculate the **P-Average**.  This primarily involves us using adjusted frequencies as opposed to unadjusted frequencies. This will be explained in the following few steps.

**Step 5: Calculate the adjusted P-Averages using the adjusted frequencies for the strata within each time period with the post-election (cases) data**

In *step 3*, we calculated the unadjusted frequency for each of the four strata as the number of days each combination occurred within the entire pre-election period divided by the total number of days in the pre-election period. We will do something similar to calculate the adjusted frequencies, but we will use post-election data instead of pre-election data. Additionally, we will calculate a P-Average for each of our 10 time periods, as opposed to calculating one frequency for the entire post-election time period.

This slide shows the frequencies (in yellow) for each time period in the post-election periods. Again, you can do this using Excel’s SUMPRODUCT function. The SUMPRODUCT function counts the number of times Nasdaq and S&P were both up, neither up, or one of them up and the other down.



We want to simulate the situation where each strata occurs equally in the pre-election as in the post-election time periods.   In order to ensure that the rate of each strata in the controls is the same as the cases, we will replace the frequency of these alternative explanations among the pre-election controls with the calculated frequency for each time period in the cases.  In this step, we are making sure that we simulate a set of controls which have the same distribution of these two stocks as each time period in the cases.

Now we can calculate the adjusted rate of Humana stock going up.  We weigh the rate with which Humana stock goes up in the controls within each strata by the new set of frequencies and will call this the *adjusted P-Averages*.  The adjusted P-Average shows the probability of Humana stock rising in the pre-election time period if it had the same frequency of occurrences of Nasdaq and S&P.



The adjusted P-Average can now be used to calculate the LCL and UCL and get the following results:



Once we plot the values above, we get the following chart:



By calculating our P-Averages with the adjusted frequencies, our modified LCL and UCL values remove the effects of the confounding variables.  Our results show us a much different picture than our previous chart. Our new analysis shows us that the 2016 Presidential election did have an effect on Humana stock, once the effect of alternative explanations were removed.  Also note that because we removed the effect of the general economy and health industry, the variation in the stock was reduced and the control limits are closer to each other.  This has allowed us to detect the effect of the election which was previously confounded with these alternative explanations.