**Test and Control**

**Contents**

1. Overview…………………………………………………………………………………………………02
2. Assumption……………………………………………………………………………………………..02
3. Objective…………………………………………………………………………………………………02
4. TNC-Tool Walk-through…………………………………………………………………………..02
   1. Input Data………………………………………………………………………………………………………02
   2. Data View……………………………………………………………………………………………………….03
   3. Project-Setup………………………………………………………………………………………………….03
   4. Time Alignment………………………………………………………………………………………………04
   5. Run Scenarios…………………………………………………………………………………………………05
      1. Month on Month Matching……………………………………………………………..05
      2. Aggregate Matching ………………………………………………………………………..06
      3. Conform Weights …………………………………………………………………………….07
   6. Results Overview…………………………………………………………………………………………….07
   7. Run Models…………………………………………………………………………………………………….10
5. **Overview**

When we want to know the efficiency of a particular event is having any effect on sales or not we do test and control analysis for that particular event.

**For Example**: If we consider promotions are given to some HCP’s they are considered as a part of Test group. The HCP’s who haven’t received that promotion are a part of control Group.

Test and control have a simple algorithm where we take the Euclidian distance between the test HCP’s and the control HCP’s and the HCPs with minimal distances are considered.

**2.0 Assumption**

Everything remains constant-It means that the test HCPs are the ones invited for the Speaker programs and the control HCP’s. Apart from this every other variable remains constant.

**3.0 Objective**

Based on the event test and control HCPs are differentiated and the sales comparison is done between them

**4.0 TNC-Tool Walk-through**

**4.1 Input Data**

Graphical user interface, text, application

Description automatically generated

* User imports the required data file in csv format. In long format year months are represented in row wise and in wide format, year months are divided into columns month wise for each variable.
* In the inputs page, user should select the required numeric, categorical, date column and format based on the imported data.
* Year month format varies-User should select based on the input file
* Download Sample file option is given where the user can see the sample dataset of wide and long format of sample dataset to get the idea of the data fomat

**4.2 Data View**

Table

Description automatically generated

* The user can view the uploaded data in Data View.
* The long format is converted into wide format.

**4.3 Project-Setup**

**Graphical user interface, text, application

Description automatically generated**

* In project Setup, project Name, Event Period, Pre-Post Start-End Dates should be specified.
* Event Date: When the event starts-Will be present only for test HCPs.

Graphical user interface, text, application

Description automatically generated

* After Project Selections, Key Identifier metrics should be selected which includes HCP Identifier (Column containing HCP Ids), TC Identifier(Column Flagged as Test and Control),Time Align Identifier(Event Variable) and Metric Identifier(Sales Variable).

**4.4 Time Alignment**

All the above explanation is before Time Alignment.

Let us consider a six-month period for 3 HCPs.

HCP A

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| June | July | August | September | October | November |

HCP B

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| June | July | August | September | October | November |

HCP C

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| June | July | August | September | October | November |

We can observe that few cells are shaded, These Shaded cells indicate the month at which HCP received the promotion.

When we calculate the distance between the two hcps, we shouldn’t blindly compare the 6-month values.

In case of HCP A and B the event started in June and the post event writings will include months June to November and pre-event writings will include 6 months before May. Similarly, if we consider HCP C the event was conducted in August. Therefore, the post event writings will include months August to January of next year, Pre event includes 6months from Feb to July.

If we want to calculate the distance between the HCP C and any other control HCP, we should make sure that we take the Post or pre-6-month period as explained above for HCP C. We should also make sure to take the same months and year we are considering for test hcp as the month and year for Control HCP.

If we are taking HCP C post event data, we consider August to January of next year. Similarly for all the control HCP whom we are considering comparing also should be from August to January of Next year.

**4.5 Run Scenarios**

**4.5.1 Month on Month Matching**

Number of Control matches for each test hcp will be specified using this slider.Chart

Description automatically generated

Graphical user interface, application, table

Description automatically generated

* In month-on-month matching, we match each month of the test id with the respective months of the control id in the pre period. Month on Month matching is usually done on variables like Details, Sales etc.
* The matching columns selected will be standardized and distance will be calculated, and number of controls ids(n)/10 for each test will be selected where distance is minimum.
* Weights can be assigned to each of the numerical variable used and distance are calculated based on the weightages assigned.

**4.5.2 Aggregate Matching**

Table

Description automatically generated

* In aggregate matching, the selected columns will be summed up for all months specified in the aggregate pre (pre\_6 is 6 months of pre period). We match unique controls which we get from MOM, and we filter number of controls ids(n)/10 for each test will be selected where distance is minimum.
* The aggregated summed up columns is standardized and used for distance calculations.
* The test and control id pairs present in both MOM and aggregated is only considered for final match and then the distances are squared.
* Weights can be assigned to each of the numerical variable used and distance are calculated based on the weightages assigned.

For better understanding of time alignment, Month on Month calculations and Aggregate calculations [TC\_Calculations](TC_Calculations_v2.xlsx).

**4.5.3 Confirm Weights**

**Text

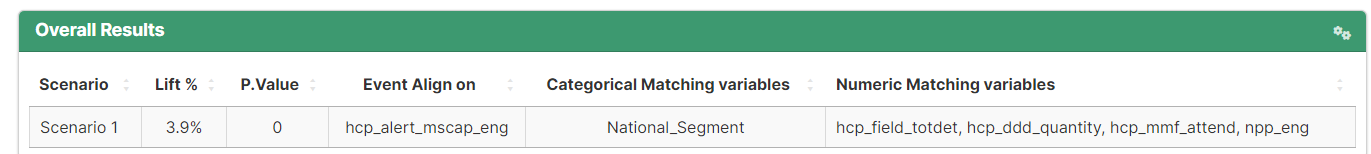
Description automatically generated**

* From the above picture, we can confirm that MOM is matching is done on sales and details on month level and aggregate matching is done on mmf and npp\_eng.
* We then launch the scenario to get the resultant matching file where each test is mapped with the best control match.

**Graphical user interface, application

Description automatically generated**

**4.6 Results Overview**

****

* In this Results Overview it shows us the lift, event align on, categorical and numeric matching variables.
* Lift indicates the percentage difference between the test and control. As from the assumption we say that everything is constant between test and control expect the event or promotion that is occurring on which analysis is performed.

Lift = Post\_Sales\_Test – Post\_Sales\_Control/Pre\_Sales\_Test.

(Or)

Lift = (Post\_Sales\_Test-Pre\_Sales\_Test) - (Post\_Sales\_Control-Pre \_Sales\_Control)

(Pre\_Sales\_Test + Pre\_sales\_Control)/2

Graphical user interface, chart, application

Description automatically generated

* This indicates No. of test ids and No. of unique matching control ids, Lift, and P value.
* The graph indicates the sales in each month, the negative x-axis points indicate pre months and positive indicates post months.
* The second graph indicate the month wise activity of all the other numeric variables.

**Graphical user interface, table

Description automatically generated**

* This shows the average table which has all the average of all the columns of both test and control to verify how close are selected control to our test.

Lift Calculations:

Lift = (Post\_Sales\_Test-Pre\_Sales\_Test) - (Post\_Sales\_Control-Pre \_Sales\_Control)

(Pre\_Sales\_Test + Pre\_sales\_Control)/2

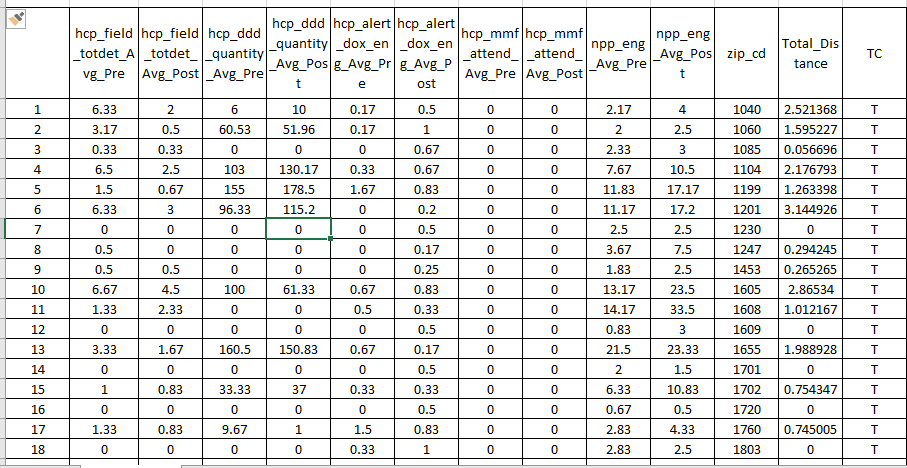
* Second table indicates about the Number of test and control ids before and after matching.
* And another table indicates Average measured metric which is sales and double difference.

**4.7 Run Models**

**Graphical user interface, text, application, chat or text message

Description automatically generated**

* In this tab we have select the dependent variable, independent variable and type of regression and click submit.



* This is the input file for our regression analysis.
* As we can see from the above table everything is in terms of average because it is not necessary that all the HCPs will have event in the particular month i.e., pre and post period will differ. so, it will be not fair if we take the total that’s why instead of summation, we consider the average of it.

**Graphical user interface

Description automatically generated with medium confidence**

* This is the model output which we get which gives us standard regression results which includes beta values, std. error, t-value, and p-value.
* We can calculate the contribution of each independent variables from its respective beta value.
* We usually considered on TCT beta value as we want to know the impact of the differentiating event of test and control.

% contribution = TCT Beta\*No. of test HCPs/Avg. sales