IOSR Journal of Mathematics (IOSR-JM)
e-ISSN: 2278-5728, p-ISSN: 2319-765X. Volume 12, Issue 3 Ver. IV (May. - Jun. 2016), PP 64-73
www.iosrjournals.org

Median Polish with Covariate on Before and After Data

Ajoge I., M.B Adam, Anwar F., and A.Y., Sadiq

Abstract: *The method of median polish with covariate is use for verifying the relationship between before and after treatment data. The relationship is base on the yield of grain crops for both before and after data in a classification of contingency table. The main effects such as grand, column and row effects were estimated using median polish algorithm. Logarithm transformation of data before applying median polish is done to obtain reasonable results as well as finding the relationship between before and after treatment data. The data to be transformed must be positive. The results of median polish with covariate were evaluated based on the computation and findings showed that the median polish with covariate indicated a relationship between before and after treatment data.*
Keywords: *Before and after data, Robustness, Exploratory data analysis, Median polish with covariate, Logarithmic transformation*

I. Introduction

This paper introduce the background of the problem by using median polish model for verifying the relationship between before and after treatment data in classification of contingency table. We first developed the algorithm of median polish, followed by sweeping of columns and rows medians. This computational procedure, as noted by Tukey (1977) operates iteratively by subtracting the median of each column from each observation in that column level, and then subtracting the median from each row from this updated table. This sequence is continuing until the median of each row and column is approximately zero (Siegel 1983).

Given a one -way or two-way median polish model with covariate would take row and column effect into consideration, which may consist of two different levels as rows and columns. If there exists an outliers or there exist a few 'bad' data values in the data, then using an iterative procedure called median polish would yield better estimates than using the mean in two-way tables Emerson & Hoaglin (1983).

The median polish is an exploratory data analysis procedure proposed by the statistician Tukey (1977) which finds an additively fit model for data in a two-way layout table of the form overall, row and column effects. Median polish has long be known and it is commonly used as a data analysis technique for examining the significance of various factors in a single or multi-way model where its algorithm is useful for removing any trends in the data by computing medians for various coordinates on the data sets (Goodall 1983). More so, median polish is a simple yet robust method to perform exploratory data analysis which is resistant to holes in the table (cells that have no values), but it may require many iterations through the data by Barbara & Wu (2003). EDA is widely used technique to determine which factors have the most influence on data values in a multi-way table or which cells in the anomalous with respect to the other cell (Shitan & Vazifedan 2012). More so, Fitrianto et al. (2014) emphasized that data can be arranged into a table, with one of the factors constant along rows and the other along columns thus denoting the observation in rows i and columns j.

Functional median polish was used in one–way and two-way model by estimating the main effects base on additive functional ANOVA model and compare the performance of traditional functional ANOVA fitted by means under different outliers models in simulation studies. A rank test was conducted to assess the significant effect of the main factors effect on weather data from climatic region effects. The result showed that functional median polish studies is robust under different outliers model while the traditional ANOVA is not resistant to outlier. (Sun & Genton 2012)

Further, median polish model technique can be used for verifying the relationship between variables. Fitrianto et al. (2014) used median polish method to verify the relationship between the students performance in relation to their attendance in a classroom. The result showed the more the attendance in the classroom, the more grades obtained. Therefore given the before and after data in the contingency table, we can use median polish model with covariate to verify the relationship and significant effect of the treatment on the crops

Logarithm transformation wass applied to the data before median polish algorithm on the contingency table to reduce interaction effect and influence of outliers. Hoaglin et al. (1983) suggests that a transformation that reduces interaction often brings other beefits by leading to residuals whose variances are more stable across different cells and whose distributions are more nearly symmetrical and as result of this, logarithm transformation therefore in the current study considered median polish incorporation of covariate to estimate the main effects. Before and after data is a process of obtaining a data from number of sites sampled prior and after the treatment Green (1979) which take place in two main sites (Control and Impact sites). Data obtained from before treatment were compared with the data after the treatment for verifying the relationship between the variables. The relationship is as a result of treatment added to the crops, that is, the more treatment added to the crops the more increased in the yields. Therefore median polish method is introduce in this research to verify the relationship between before and after the treatment on the data sets

1.1 Concept of Median polish

The relationship between the rate (response variable) and the two factors in the table (site and periods) can be expressed using a simple model z_{ij} +k_{ij} where z_{ij} is the original value, k_{ij} is the estimated value from model, α is the grand effect (common value for the whole table), τ_i is the row effect of row i, δ_j is the column effect of column and e_{ij} is the residual while i = 1,… , n. j = 1, …, k. [4].

For such data, one needs to understand the way in which the variable Z depends on the two factors. When the row and column factors are associated with numerical values g_i andg_j, the simplest description would be a linear regression with fitted cell values,

$k_{ij} = b_0 + b_1$gi +$b_2$gj (1)

The observations, z_{ij} is written as sum of two terms

$z_{ij} = k_{ij} + e_{ij}$

That is, the fit plus residuals. If no numerical values of g_i and g_j are at hand to apply the linear regression, one can still use this same idea by replacing the b_0 with a grand effect α, $b_1 g_i$ with a row effect τ_i and the $b_2 g_j$ with column effect δj , which leads to

z_{ij} = α +τ_i+ δ_j (2)

and it can be transformed into k_{ij} = α + τ_i+ δ_j +e_{ij} (3)

Where z_{ij} = observation from row i and column j, α is overall effect, τ_i is i^{th} row effect, and δ_j is the j^{th} column effect.

II. Algorithm of using median polish

Median polish algorithm works by alternately removing the row and column medians, and continues until the last number of iterations specified (Siegel 1983). In principle, the process continues until the rows and columns median have zero median.

1. Take the median of each row and record the value to the side of the row, subtract the row median from each value in that row.
2. Calculate the median of the row medians, and record the value as the overall effect, subtract the overall effect from each of the row medians.
3. Take the median of each column and record the value beneath the column. Subtract the column median from each value in that particular column.
4. Calculate the median of the column medians and add the value to the current overall effect. Subtract this addition to the overall effect from each of the column medians.
5. Repeat steps 1 – 4 until no changes occur with the row or column medians.

2.1 One-way median polish model with covariate

The variable x_{ij} is for before treatment data, which serve as a covariate in the model. As a covariate, it has some interaction effect or association with z_{ij} . The variable z_{ij} is after treatment data. With covariate, one-way model becomes:

z_{ij} = α +δ_j+φx_{ij} + e_{ij}

Where i = 1,• • •,n , j = 1,• • •, k and α represents the grand effect and δ_j is the column effect and e_{ij} is the residual while φ is the coefficient of the covariate x_{ij}

2.2 Two-way median polish model with covariate

Unlike one-way model with covariate, two-way model with covariate would take row effect into consideration which may consist of two different levels as rows 1,…, n and columns 1,…, k. We start the process beginning from the column levels first before in the row levels.

With covariate, two-way model can be written as:

z_{ij} = α + τ_i+ δ_j+ φ x_{ij} + e_{ij}

where i = 1,• • •,n , j = 1,• • •,k and α is the grand effect, τ_i is the row effect, δ_j is the column effect and e_{ij} is the residual while φ is the coefficient of the covariate x_{ij}

This emphasizes the use of medians in every stage of the fitting process in order to summarize either rows or columns, and sweeping the information, they describe into the fit. The common term describes the level of the data values in the table as a whole and can describe a two-way table that has the same constant value in each cell. However, each row effect describes the way in which the data values in its row tend to differ from the common level [2]. Similarly, the column effects describe the way in which the data values in each column tends to differ from the common value.

III. Stopping criterion

This is the stopping stage for the median polish with covariate, where the procedures have reached a final stage after a series of iterations and usually to the final stage is when the median for row and column effects is approximately zero (Siegel 1983). In principle, the median polish process continues by repeated extraction of row and column until convergence with respect to a stopping criterion. In practice, the process stops when it gets sufficiently close to that goal where the sweeping operation can no more iterate again.

IV. Description of Data

We took an empirical data from ministry of Agriculture and farm extension, Adamawa state of Nigeria. There are total of 10 locations/sites that we used as sample to run the median polish algorithm. All the 10 sites can be grouped into 12 categories; Jan- Dec. The description of the data is stated here for the obvious reason of being able to follow this write -up. Much description of the data including all outcome variables, which have been in its original form. A randomized block design is used with 12 columns as the periods (Jan.-Dec) and10 rows as the sites (1-10) were displayed in a contingency table of n-row and k-column vectors respectively. The studied variable was grain yields (kg/bags) and for this work yields for each cell on each of the locations comprised of data values. Therefore, we are going to run median polish method based on data from 10 sites which were used as sample to run the median polish algorithm that was grouped into 12 categories (Jan- Dec.) as stated above.

V. Methodology

An independent variable x_{ij} (known as a covariate) plays a vital role in the model. As a covariate, it has relationship or association with z_{ij} (Shitan & Vazifedan 2012). The x_{ij} is for before treatment data while z_{ij} is after treatment data. With covariate two-way model can be written as

$z_{ij} = \alpha + \tau_i + \delta_j + \varphi\, x_{ij} + e_{ij}$ where i = 1,• • • ,n, j = 1,• • •,k, and α is the overall effect, τ_i is the row effect, δ_j is the column effect, e_{ij} is the residual while φ = b is the coefficient of the covariate.

The technique of fitting a model with covariate, needs an initial fitting of

$z_{ij} = \bar{\alpha} + \bar{\tau}_i + \bar{\delta}_j + \bar{e}_{ij}$ and $x_{ij} = \alpha^* + \tau_i^* + \delta_j^* + e_{ij}^*$

.

Note that ($\bar{e}_{ij}$) is the final residuals of after data while (e_{ij}^*) is for the final residuals of table of before data for two-way model. Then we fit resistant line of residuals of $\bar{e}_{ij}$ against residuals of e_{ij}^*.

$$\bar{e}_{ij} = \alpha + b\, e_{ij}^* \quad (4)$$

$$z_{ij} = \bar{\alpha} + \bar{\tau}_i + \bar{\delta}_j + \bar{e}_{ij} \quad (5)$$

$$\bar{e}_{ij} = z_{ij} - \bar{\alpha} - \bar{\tau}_i - \bar{\delta}_j \quad (6)$$

$$z_{ij} - \bar{\alpha} - \bar{\tau}_i - \bar{\delta}_j = a + b e_{ij}^* \quad (7)$$

$$z_{ij} - \bar{\alpha} - \bar{\tau}_i - \bar{\delta}_j = a + b\,(x_{ij} - \alpha^* - \tau_i^* - \delta_j^*) \quad (8)$$

$$z_{ij} - \bar{\alpha} - \bar{\tau}_i - \bar{\delta}_j = a + b x_{ij} - b\alpha^* - b\tau_i^* - b\,\delta_j^* \quad (9)$$

$$z_{ij} = \bar{\alpha} + \bar{\tau}_i + \bar{\delta}_j + \alpha + b x_{ij} - b\alpha^* - b\tau_i^* - b\,\delta_j^* \quad (10)$$

$$z_{ij} = (\bar{\alpha} + \alpha - b\alpha^*) + (b\bar{\tau}_i - b - b\tau_i^*) + (\bar{\delta}_j - b\,\delta_j^*) + b x_{ij} \quad (11)$$

Then from equation (11)

z_{ij} is the response variable (i,j) where, i = 1,• • •,n, j = 1,• • •,k

($\bar{\alpha}$+ a − bα^*) is the overall effect

(b$\bar{\tau}_i$− bτ_i^*) is the row effect

($\bar{\delta}_j$ − b δ_j^*) is the column effect

φ = b is the coefficient of the covariate while x_{ij} is the covariate. Therefore to fit resistant line of the residuals $\bar{e}_{ij}$ against e_{ij}^* we need to sort the data in ascending or descending order by a variable and divide the batch into three parts; Left, middle and right respectively.

Note that number of residuals points along the batch may not exactly divisible by three. Take for instance, when n is divided into 3, the remainder could be 1 or 2. if the remainder is 1, the extra point is allocated to the

middle portion but if there are 2 extra points then one of them should be allocated to the left part while the remaining one should be allocated to the right part so that we can obtain the value for ”a” and ”b” (Shitan & Vazefidan 2012).
In this write-up the value for “a” and “b” is obtained by finding the residuals from computer using R-programme.

VI. Result & Discussion

6.1 Result

Here we generate a two-way table, which include before and after treatment data in respective locations. To fulfill the main objectives that is set in this study, step by step of the median polish algorithm is done by using R- programming . After reaching final residuals table, which includes row effects, column effects and overall effect, then we analyze the results. Analysis of before and after treatment data is shown on the tables as follows:

6.1.1 Two-way model for the before treatment data

Table 1: Before treatment data

Site	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec
1	133	122	115	122	113	124	110	130	107	110	103	118
2	115	102	101	113	103	102	117	107	118	100	119	104
3	120	123	125	102	122	107	114	100	113	103	135	105
4	120	100	115	101	118	115	107	106	106	101	100	100
5	100	121	114	111	116	107	111	106	113	100	102	111
6	114	124	102	116	100	108	107	132	108	105	100	110
7	113	128	100	107	112	106	115	117	123	112	108	110
8	118	120	129	100	113	114	103	120	110	104	115	113
9	128	100	101	100	107	102	113	115	102	118	107	104
10	103	108	107	104	122	110	102	114	104	106	124	102

Table 1 shows raw data of before treatment data of the grains yields obtained on a contingency table of n-row and k-column vectors respectively.

Table 2: Logarithm transformation of before treatment data with the column medians

Site	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec
1	4.890	4.804	4.745	4.804	4.727	4.820	4.701	4.868	4.673	4.701	4.635	4.771
2	4.745	4.625	4.615	4.727	4.635	4.625	4.762	4.673	4.771	4.605	4.779	4.644
3	4.788	4.812	4.828	4.625	4.804	4.673	4.736	4.605	4.727	4.635	4.905	4.664
4	4.787	4.605	4.745	4.615	4.771	4.745	4.673	4.663	4.663	4.615	4.605	4.605
5	4.605	4.796	4.736	4.710	4.754	4.673	4.710	4.664	4.727	4.605	4.625	4.710
6	4.736	4.820	4.625	4.754	4.605	4.682	4.673	4.883	4.682	4.654	4.605	4.701
7	4.727	4.852	4.605	4.673	4.719	4.664	4.745	4.762	4.812	4.719	4.682	4.701
8	4.771	4.788	4.860	4.605	4.727	4.736	4.634	4.788	4.701	4.644	4.745	4.727
9	4.635	4.605	4.615	4.605	4.673	4.625	4.727	4.745	4.625	4.771	4.673	4.644
0	4.635	4.682	4.673	4.644	4.804	4.701	4.625	4.736	4.644	4.663	4.820	4.625
cm	4.741	4.792	4.705	4.659	4.727	4.67	4.705	4.741	4.691	4.649	4.678	4.682

Table 2 shows the log transformation of before treatment data of Table 1 along with column medians.

Table 3: Difference between data in Table 2 with its column medians

Site	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec
1	0.150	0.012	0.040	0.145	0.000	0.143	-0.005	0.127	-0.019	0.051	-0.043	0.089
2	0.004	-0.167	-0.089	0.069	-0.093	-0.053	0.057	-0.068	0.079	-0.044	0.102	-0.038
3	0.047	0.021	0.124	-0.034	0.077	-0.005	0.031	-0.135	0.036	-0.015	0.228	-0.019
4	0.047	-0.187	0.040	-0.044	0.043	0.068	-0.032	-0.077	-0.028	-0.034	-0.072	-0.077
5	-0.135	0.004	0.032	0.051	0.026	-0.005	0.005	-0.077	0.036	-0.044	-0.053	0.028
6	-0.004	0.029	-0.080	0.095	-0.122	0.005	-0.032	0.142	-0.009	0.005	-0.072	0.019
7	-0.013	0.060	-0.099	0.014	-0.009	-0.014	0.040	0.022	0.121	0.069	0.005	0.019
8	0.030	-0.004	0.155	-0.054	0.000	0.059	-0.070	0.047	0.009	-0.005	0.068	0.045
9	-0.106	-0.187	-0.089	-0.053	-0.055	-0.053	0.022	0.004	-0.066	0.122	-0.005	-0.038
0	-0.106	-0.110	-0.032	-0.014	0.077	0.023	-0.080	-0.004	-0.047	0.014	0.143	-0.057
cm	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000

Table 3 display deviations which are the outcomes of subtraction from the values from Table 2 with respective levels of columns medians. Finally, the last row shows the column medians of the respective column levels

Table 4: Row medians & overall effect of the data

Site	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec	R.M.
1	0.150	0.012	0.040	0.145	0.000	0.143	-0.005	0.127	-0.019	0.051	-0.043	0.089	0.046
2	0.004	-0.167	-0.089	0.069	-0.093	-0.053	0.057	-0.068	0.079	-0.044	0.102	-0.038	-0.041
3	0.047	0.021	0.124	-0.034	0.077	-0.005	0.031	-0.135	0.036	-0.015	0.228	-0.019	0.026
4	0.047	-0.187	0.040	-0.044	0.043	0.068	-0.032	-0.077	-0.028	-0.034	-0.072	-0.077	-0.033
5	-0.135	0.004	0.032	0.051	0.026	-0.005	0.005	-0.077	0.036	-0.044	-0.053	0.028	0.004
6	-0.004	0.029	-0.080	0.095	-0.122	0.005	-0.032	0.142	-0.009	0.005	-0.072	0.019	0.000
7	-0.013	0.060	-0.099	0.014	-0.009	-0.014	0.040	0.022	0.121	0.069	0.005	0.019	0.016
8	0.030	-0.004	0.155	-0.054	0.000	0.059	-0.070	0.047	0.009	-0.005	0.068	0.045	0.020
9	-0.106	-0.187	-0.089	-0.053	-0.055	-0.053	0.022	0.004	-0.066	0.122	-0.005	-0.038	-0.053
0	-0.106	-0.110	-0.032	-0.014	0.077	0.023	-0.080	-0.004	-0.047	0.014	0.143	-0.057	-0.023
cm	4.741	4.792	4.705	4.659	4.727	4.675	4.705	4.741	4.691	4.649	4.678	4.682	4.698

The last row of Table 4 shows the column medians for the respective levels of column medians. We search across each row to obtain the row medians at the last column and we search the medians column in the last row to obtain the overall effect (4.698) for all column medians.

Table 5: Difference between the data in Table 4 & its row medians

Site	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec	Rm
1	0.104	-0.034	-0.005	0.100	-0.046	0.097	-0.050	0.081	-0.064	0.005	-0.089	0.043	0.000
2	0.045	-0.126	-0.049	0.110	-0.052	-0.012	0.098	-0.027	0.120	-0.003	0.142	0.003	0.000
3	0.021	-0.005	0.098	-0.060	0.051	-0.031	0.005	-0.161	0.010	-0.040	0.202	-0.044	0.000
4	0.080	-0.153	0.074	-0.010	0.076	0.101	0.001	-0.044	0.005	-0.001	-0.039	-0.044	0.000
5	-0.140	-0.000	0.027	0.047	0.022	-0.009	0.000	-0.082	0.032	-0.049	-0.057	0.023	0.000
6	-0.005	0.029	-0.080	0.095	-0.122	0.005	-0.032	0.142	-0.009	0.005	-0.073	0.018	0.000
7	-0.030	0.044	-0.116	-0.002	-0.025	-0.030	0.024	0.005	0.105	0.053	-0.012	0.002	0.000
8	0.011	-0.024	0.136	-0.073	-0.020	0.039	-0.090	0.027	-0.011	-0.024	0.048	0.026	0.000
9	-0.053	-0.133	-0.036	-0.001	-0.002	0.001	0.076	0.057	-0.013	0.175	0.048	0.015	0.000
1	-0.083	-0.087	-0.009	0.009	0.100	0.046	-0.057	0.019	-0.024	0.037	0.166	-0.034	0.000
cm	4.741	4.792	4.705	4.659	4.727	4.675	4.705	4.741	4.691	4.649	4.678	4.682	4.698

The values in Table 5 are the outcome of subtraction of values from Table 4 with the respective levels of row medians. The Last row of the shows column medians of the observations whiles the overall effect =4.698 is display at the extreme end of the last row

Table 6: The final residuals, row and column effects

Site	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec	R.E
1	0.104	-0.034	-0.005	0.100	-0.046	0.097	-0.050	0.081	-0.064	0.005	-0.089	0.043	0.046
2	0.045	-0.126	-0.049	0.110	-0.052	-0.012	0.098	-0.027	0.120	-0.003	0.142	0.003	-0.041
3	0.021	-0.005	0.098	-0.060	0.051	-0.031	0.005	-0.161	0.010	-0.040	0.202	-0.044	0.026
4	0.080	-0.153	0.074	-0.010	0.076	0.101	0.001	-0.044	0.005	-0.001	-0.039	-0.044	-0.033
5	-0.140	-0.000	0.027	0.047	0.022	-0.009	0.000	-0.082	0.032	-0.049	-0.057	0.023	0.044
6	-0.005	0.029	-0.080	0.095	-0.122	0.005	-0.032	0.142	-0.009	0.005	-0.073	0.018	0.000
7	-0.030	0.044	-0.116	-0.002	-0.025	-0.030	0.024	0.005	0.105	0.053	-0.012	0.002	0.016
8	0.011	-0.024	0.136	-0.073	-0.020	0.039	-0.090	0.027	-0.011	-0.024	0.048	0.026	0.020
9	-0.053	-0.133	-0.036	-0.001	-0.002	0.001	0.076	0.057	-0.013	0.175	0.048	0.015	-0.053
10	-0.083	-0.087	-0.009	0.009	0.100	0.046	-0.057	0.019	-0.024	0.037	0.166	-0.034	-.023
C.E.	0.043	0.094	0.007	-0.039	0.029	-0.023	0.007	0.043	-0.007	-0.049	-0.020	-0.017	0.000

Table 6 displays the final residuals, row and column effects. The column medians each subtracted from overall effect (4.698) to get a treatment (column) effect which written on the last row.

Now the process of sweeping is complete for two-way model for before treatment data. This shows that it has reached stopping stage where the median polish procedures have reached a final stage that is when the median for row and column median is zero.

Hence, we have the model to be;

Z_{IJ} = 4.698 + τ_I+ δ_J+e_{ij}

where,

Z_{ij} is the response variable, α is the overall effect=4.698, τ_i is the row effect, δ_j is treatment effect and e_{ij} are the residuals

6.1.2 Two-way model for after treatment data

Table 7: After treatment data on the contingency table

Site	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec
1	401	394	381	374	371	404	367	363	382	391	397	365
2	386	385	376	396	354	388	394	369	380	359	383	385
3	376	361	406	398	391	396	352	355	387	383	375	384
4	387	357	364	388	364	368	406	370	374	391	395	361
5	392	368	375	397	362	358	370	357	383	365	359	362
6	393	387	386	418	366	394	405	378	380	382	395	395
7	404	368	391	350	392	376	363	411	380	374	390	390
8	401	413	381	382	415	411	369	393	388	413	380	394
9	381	407	358	364	386	363	377	384	379	386	387	367
10	384	369	388	363	391	405	409	371	388	356	404	380
md	389.5	377	381	385	378.5	391	373.5	370.5	381	382.5	388.5	382

Table 7 shows the after treatment data in the contingency table

Table 8: Logarithm transformation of after treatment data with the column medians

Site	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec
1	5.994	5.976	4.943	5.924	5.916	6.002	5.905	5.894	5.945	5.969	5.984	5.900
2	5.956	5.953	5.930	5.982	5.869	5.961	5.976	5.911	5.940	5.883	5.948	5.953
3	5.930	5.889	6.006	5.987	5.969	5.981	5.864	5.872	5.958	5.948	5.927	5.951
4	5.958	5.878	5.897	5.961	5.897	5.908	6.006	5.914	5.924	5.969	5.979	5.889
5	5.971	5.908	5.927	5.984	5.892	5.881	5.914	5.878	5.948	5.900	5.883	5.892
6	5.974	5.958	5.956	6.036	5.903	5.976	6.004	5.935	5.940	5.945	5.979	5.979
7	6.001	5.908	5.969	5.858	5.971	5.929	5.894	6.019	5.940	5.924	5.966	5.966
8	5.994	6.024	5.943	5.945	6.028	6.019	5.911	5.974	5.961	6.024	5.940	5.976
9	5.943	6.009	5.881	5.897	5.956	5.894	5.932	5.951	5.938	5.956	5.958	5.905
10	5.951	5.911	5.961	5.894	5.969	6.004	6.014	5.916	5.961	5.875	6.001	5.940
cm	5.965	5.932	5.943	5.953	5.936	5.969	5,923	5.915	5.943	5.947	5.962	5.945

Table 8 shows the log-transformed data of Table 7 with the column medians

Table 9: Difference between data in Table 8 and its column medians

Site	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec
1	0.029	0.044	0.000	-0.029	-0.020	0.033	-0.018	-0.021	0.003	0.022	0.022	-0.046
2	-0.009	0.021	-0.013	0.028	-0.067	-0.008	0.054	-0.004	-0.003	-0.063	-0.014	0.008
3	-0.035	-0.043	0.064	0.033	0.033	0.013	-0.059	-0.043	0.016	0.001	-0.035	0.005
4	-0.006	-0.054	-0.046	0.008	-0.039	-0.061	0.084	-0.001	-0.019	0.022	0.017	-0.057
5	0.006	-0.024	-0.016	0.031	-0.044	-0.088	-0.009	-0.037	0.005	-0.047	-0.079	-0.054
6	0.009	0.026	0.013	0.082	-0.033	0.008	0.081	0.020	-0.003	-0.001	0.017	0.034
7	0.037	-0.024	0.026	.-0.095	0.035	-0.039	-0.029	0.104	-0.003	-0.023	0.004	0.021
8	0.029	0.092	0.000	-0.008	0.092	0.050	-0.012	0.059	0.018	0.077	-0.022	0.031
9	-0.022	0.077	-0.062	-0.056	0.020	-0.074	0.009	0.036	-0.005	0.009	-0.004	-0.040
10	-0.014	-0.021	0.018	-0.059	0.033	0.035	0.091	0.001	0.018	-0.072	0.039	-0.005
cm	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000

Table 9 shows the deviations, which are the outcomes of subtraction from the values from Table 8 with respective levels of column medians. Finally, the last row shows the column medians of the respective column levels.

Table 10: Row medians and overall effect of the data

Site	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec	rm
1	0.029	0.044	0.000	-0.029	-0.020	0.033	-0.018	-0.021	0.003	0.022	0.022	-0.046	0.001
2	-0.009	0.021	-0.013	0.028	-0.067	-0.008	0.054	-0.004	-0.003	-0.063	-0.014	0.008	-0.006
3	-0.035	-0.043	0.064	0.033	0.033	0.013	-0.059	-0.043	0.016	0.001	-0.035	0.005	0.003
4	-0.006	-0.054	-0.046	0.008	-0.039	-0.061	0.084	-0.001	-0.019	0.022	0.017	-0.057	-0.013
5	0.006	-0.024	-0.016	0.031	-0.044	-0.088	-0.009	-0.037	0.005	-0.047	-0.079	-0.054	-0.031
6	0.009	0.026	0.013	0.082	-0.033	0.008	0.081	0.020	-0.003	-0.001	0.017	0.034	0.015
7	0.037	-0.024	0.026	.-0.095	0.035	-0.039	-0.029	0.104	-0.003	-0.023	0.004	0.021	0.001
8	0.029	0.092	0.000	-0.008	0.092	0.050	-0.012	0.059	0.018	0.077	-0.022	0.031	0.030
9	-0.022	0.077	-0.062	-0.056	0.020	-0.074	0.009	0.036	-0.005	0.009	-0.004	-0.040	-0.005
1	-0.014	-0.021	0.018	-0.059	0.033	0.035	0.091	0.001	0.018	-0.072	0.039	-0.005	0.010
cm	5.965	5.932	5.943	5.953	5.936	5.969	5,923	5.915	5.943	5.947	5.962	5.945	5.944

The values in Table 10 are the outcome of subtraction of values from Table 9 with the respective levels of column medians. We search across each row to obtain the row medians at the last column and we search the medians column in the last row to obtain the overall effect (5.944) for all column medians.

Table 11: Final residuals with column and row effects

Site	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec	RE
1	0.028	0.043	-0.001	-0.030	-0.021	40.031	-0.019	-0.022	0.001	0.021	0.020	-047	0.001
2	-0.003	0.027	-0.007	0.034	-0.061	-0.002	0.059	0.002	0.003	-0.058	-0.008	0.014	-0.006
3	-0.039	-0.046	0.060	0.030	0.029	0.010	-0.063	-0.046	0.012	-0.002	-0.039	0.002	0.003
4	0.006	-0.042	-0.033	0.020	-0.026	-0.048	0.096	0.011	-0.006	0.035	0.029	-0.044	-0.013
5	0.037	0.007	0.015	0.061	-0.014	-0.058	0.021	-0.007	0.036	-0.016	-0.048	-0.023	-0.031
6	-0.006	0.012	-0.002	0.068	-0.048	-0.007	0.066	0.005	-0.017	-0.016	0.002	.019	0.015
7	0.036	-0.025	0.025	-0.096	0.035	-0.040	-0.029	0.103	-0.003	-0.023	0.003	0.020	0.001
8	-0.001	0.061	-0.030	-0.038	0.062	0.020	-0.042	0.029	-0.012	0.047	-0.052	0.001	0.030
9	-0.018	0.081	-0.058	-0.052	0.024	-0.070	0.014	0.040	-0.001	0.014	0.001	-0.036	-0.005
10	-0.024	-0.031	0.008	-0.069	0.023	0.025	0.081	-0.008	0.008	-0.082	0.029	-0.015	0.010
C.E.	0.021	-0.012	-0.001	-0.009	-0.008	0.025	-0.021	-0.029	-0.001	0.003	0.018	0.001	0.000

Table 11 shows the final table, which shows the residuals and column effect and row effect. The column medians each subtracted from overall effect (5.944) to get a new treatment effect which written on the last row.

The process of sweeping is complete in two-way table for after data, this is the last stage of process of the two-way median polish. This shows that it has reached stopping stage where the median polish procedures have reached a final stage that is when the median for row and column effect is approximately zero.

To fit resistant line of residuals of $\bar{e}_{ij}$ against residuals of e^*_{ij}

We can now use ($\bar{e}_{ij}$) as the final residuals for after treatment data in Table 11 and (e^*_{ij}) as the final residuals of before data in Table 6 of two-way model.

Table 12: Final residuals, row and column effects of before data (e^*_{ij})

Site	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec	R.E
1	0.104	-0.034	-0.005	0.100	-0.046	0.097	-0.050	0.081	-0.064	0.005	-0.089	0.043	0.046
2	0.045	-0.126	-0.049	0.110	-0.052	-0.012	0.098	-0.027	0.120	-0.003	0.142	0.003	-0.041
3	0.021	-0.005	0.098	-0.060	0.051	-0.031	0.005	-0.161	0.010	-0.040	0.202	-0.044	0.026
4	0.080	-0.153	0.074	-0.010	0.076	0.101	0.001	-0.044	0.005	-0.001	-0.039	-0.044	-0.033
5	-0.140	-0.000	0.027	0.047	0.022	-0.009	0.000	-0.082	0.032	-0.049	-0.057	0.023	0.044
6	-0.005	0.029	-0.080	0.095	-0.122	0.005	-0.032	0.142	-0.009	0.005	-0.073	0.018	0.000
7	-0.030	0.044	-0.116	-0.002	-0.025	-0.030	0.024	0.005	0.105	0.053	-0.012	0.002	0.016
8	0.011	-0.024	0.136	-0.073	-0.020	0.039	-0.090	0.027	-0.011	-0.024	0.048	0.026	0.020
9	-0.053	-0.133	-0.036	-0.001	-0.002	0.001	0.076	0.057	-0.013	0.175	0.048	0.015	-0.053
10	-0.083	-0.087	-0.009	0.009	0.100	0.046	-0.057	0.019	-0.024	0.037	0.166	-0.034	-.023
C.E.	0.043	0.094	0.007	-0.039	0.029	-0.023	0.007	0.043	-0.007	-0.049	-0.020	-0.017	4.698

Table 12 shows the residuals, row and column effects of before treatment data

Table 13: Final residuals, column and row effects of after treatment data ($\bar{e}_{ij}$)

Site	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec	RE
1	0.028	0.043	-0.001	-0.030	-0.021	40.031	-0.019	-0.022	0.001	0.021	0.020	-047	0.001
2	-0.003	0.027	-0.007	0.034	-0.061	-0.002	0.059	0.002	0.003	-0.058	-0.008	0.014	-0.006
3	-0.039	-0.046	0.060	0.030	0.029	0.010	-0.063	-0.046	0.012	-0.002	-0.039	0.002	0.003
4	0.006	-0.042	-0.033	0.020	-0.026	-0.048	0.096	0.011	-0.006	0.035	0.029	-0.044	-0.013
5	0.037	0.007	0.015	0.061	-0.014	-0.058	0.021	-0.007	0.036	-0.016	-0.048	-0.023	-0.031
6	-0.006	0.012	-0.002	0.068	-0.048	-0.007	0.066	0.005	-0.017	-0.016	0.002	.019	0.015
7	0.036	-0.025	0.025	-0.096	0.035	-0.040	-0.029	0.103	-0.003	-0.023	0.003	0.020	0.001
8	-0.001	0.061	-0.030	-0.038	0.062	0.020	-0.042	0.029	-0.012	0.047	-0.052	0.001	0.030
9	-0.018	0.081	-0.058	-0.052	0.024	-0.070	0.014	0.040	-0.001	0.014	0.001	-0.036	-0.005
10	-0.024	-0.031	0.008	-0.069	0.023	0.025	0.081	-0.008	0.008	-0.082	0.029	-0.015	0.010
C.E.	0.021	-0.012	-0.001	-0.009	-0.008	0.025	-0.021	-0.029	-0.001	0.003	0.018	0.001	5.944

The values in Table 13 are the final residuals of after treatment data in Table 11 with rows and column effects. The value for "a" and "b" is obtained by finding the residuals from computer using R- programme as

$$b = \frac{e^*_{R-}e^*_L}{\bar{e}_{R-e_L}} = 0.0586 \qquad (14)$$

$$\alpha = 1/3[(\bar{e}_L + \bar{e}_M + \bar{e}_R) - b(e_L^* + e_M^* + e_R^*)] = 0.0003 \quad (15)$$

Therefore,
The overall effect is

$$\alpha = \bar{\alpha} + a - b\alpha^* \quad (16)$$

α=5.944+0.0003-0.0586(4.698) =5.669

The row effect

$$\tau_i = \bar{\tau}_i\tau - b\tau^* \quad (17)$$

Table 14: Estimated row effects

S/N	$\bar{\tau}_i$	τ^*	$b\tau_i^*$	$\tau_i = b\tau_i - b\tau^*\tau^*$
1	0.001	0.046	0.003	-0.002
2	-0.006	-0.041	-0.002	-0.004
3	0.003	0.026	0.002	0.002
4	-0.013	-0.033	-0.019	-0.011
5	-0.031	0.004	0.000	-0.031
6	0.015	0.000	0.000	0.015
7	0.001	0.016	0.001	0.000
8	0.030	0.020	0.001	0.029
9	-0.005	-0.053	-0.003	-0.002
10	0.010	-0.023	-0.001	0.011

Table 15: Estimated treatment effects

S/N	$\bar{\delta}_j$	δ_j^*	$b\delta_j^*$	$\delta_j = \delta_j - b\delta_j^*$
1	0.021	0.043	0.003	0.021
2	-0.012	0.094	0.006	-0.018
3	-0.001	0.007	0.001	-0.001
4	0.009	-0.039	-0.002	0.011
5	-0.008	0.029	0.002	-0.010
6	0.025	-0.023	-0.001	0.026
7	-0.021	0.007	0.000	-0.021
8	-0.029	0.043	0.003	-0.032
9	-0.001	-0.007	-0.000	-0.001
10	0.003	-0.049	-0.003	0.006
11	0.018	-0.020	-0.001	0.019
12	0.001	-0.017	-0.001	0.002

and the treatment effect

$$\delta_j = \bar{\delta}_j - b\delta_j^* \quad (18)$$

The estimated residuals can be obtain by

$$e_{ij} = \bar{e}_{ij} - a + be_{ij}^* \quad (19)$$

Table 16: Summary of estimated residuals

S/N	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec
1	0.022	0.050	-0.001	-0.036	-0.023	0.025	-0.023	-0.027	0.005	0.020	0.025	-0.050
2	-0.006	0.034	-0.009	0.027	-0.058	-0.002	0.053	0.003	-0.004	-0.058	-0.017	0.014
3	-0.041	-0.046	0.063	0.023	0.026	0.010	-0.064	-0.037	0.011	-0.004	-0.051	0.004
4	0.001	-0.051	-0.028	0.020	-0.031	-0.054	0.096	0.013	-0.007	0.035	0.031	-0.046
5	0.045	0.007	0.013	0.058	-0.016	-0.058	0.021	-0.003	0.034	-0.013	-0.045	-0.025
6	-0.006	0.010	0.002	0.062	-0.041	-0.008	0.068	0.004	-0.017	-0.017	0.006	0.018
7	0.038	-0.028	0.026	-0.101	0.036	-0.039	-0.031	0.103	-0.009	-0.026	0.003	0.020
8	-0.002	0.0.062	-0.038	-0.034	0.063	0.017	-0.037	0.027	-0.012	0.048	-0.055	0.001
9	-0.015	0.089	-0.056	-0.050	0.024	-0.070	0.009	0.036	-0.001	0.003	-0.002	-0.037
10	-0.019	-0.026	0.008	-0.070	0.017	0.022	0.084	-0.009	0.009	-0.085	-0.039	0.013

Hence, the model becomes;

$$z_{ij} = \alpha + \tau_i + \delta_j + \varphi x_{ij} + e_{ij} \quad (20)$$

where z_{ij} = mean output from row i and column j, α is overall effect, τ_i is i^{th} row effect, δ_j is j^{th} column effect and e_{ij} is the residuals while i = 1,· · · ,n, j = 1,· · · ,k and φ is the coefficient of the covariate with x_{ij} as the covariate that shows the relationship between before and after the data .

Hence, the model can be written as:

$$z_{ij} = 5.944 + \tau_i + \delta_j + 0.0586x_{ij} + e_{ij} \quad (21)$$

where,

$\alpha = (\bar{\alpha} + a - b\alpha^*)$ =5.944 is the overall effect,

$\tau_i = \bar{\tau}_i - b\tau_i^*$ is the row effect,

$\delta_j = (\bar{\delta}_j - 0.0586\delta^*)$ is the treatment effect

φ=0.0586 is the coefficient of the covariate

x_{ij} is independent variable that serve as covariate

and e_{ij} is the residuals

This is the stopping stage for two-way median polish model with covariate. It is the last stage where the median polish procedures has reached a after a series of iterations

VII. Discussion

Based on the result obtained from two-way median polish with covariate, we can able to estimate overall, row, treatment effects and residuals from equations (16) -(19). We can predict the response variable using equation 21.

$Z_{11} = 5.669 + -.002 + 0.021 + (0.0586)(4.890) + 0.022 =$	(22)
$Z_{11} = 5.669 + -.002 + 0.021 + 0.283554 + 0.022 = 5.994$	(23)

and so on. We can see that covariate has brought the relation/ association between before and after data.

By looking at the respective function, we apply the equation (20) to analyze the data

Suppose we want to get back the observation from row 7 and column 6 in Table 12, which is Z_{76} . Then we estimate using 5.669 + 0.000 + 0.026 + (0.0586)(4.664) − 0.039 = 5.929 which is the same value of Z_{76} in Table 8.

Now, we shall look at estimated treatment effect of Table 15, which might be able to bring us to a clear relationship between before and after data. The Jan. yields is 0.021- -0.018= 0.039=3.9 percent points better (on average) than the Feb. yields, Feb. yields is (-0.018- -0.001= - 0.017)= 1.7 percent point lower than March harvest. Then April group have (0.011- -0.001=0.012)=1.2 percent higher point than the month of March and so on.

The same process applicable to estimated row effect on Table 14.

VIII. Conclusion

In this paper, we have described an appropriate method of exploratory data analysis (EDA) using median polish with covariate to verify the relationship between before and after treatment data. We have shown that the method scale with covariate allowing us to perform EDA in finding the relationship between before and after treatment data. The main effects such as grand, row and column effects were equally estimated using median polish algorithm. The model of median polish with covariate can be further tested using IQRoQ plots to measure the spread of rows and columns effects

References

[1]. Emerson, J. D. & Hoaglin, D. C. (1983).Analysis of two-way tables by medians. Understanding Robust and Exploratory Data Analysis 165–210.

[2]. Fitrianto A. et al. (2014). Median Polish for final Grades of MTH3000 and MTH 4000- Level courses. .Journal of Applied of Mathematical Science Vol. 3:126, 6295-6302

[3]. Freeman, G.H. (1973). Analysis of Interaction in Incomplete Two-way Tables, National Vegetable Research Station, Wellesbourne, Warwick, 1973 Green, R.H. (1979). Sampling design &statistical methods for environmental biologists.

[4]. John Wiley & Sons Goodall C. (1983). Examining Residuals; Understanding Robust and Exploratory Data Analysis, New York: John Wiley & Sons 211-246

[5]. Hoaglin, D.C., Mosteller F. Tukey, J.W. (1983). Understanding robust and exploratory data analysis. New York: John Wiley & Sons,
[6]. Hoaglin, D.C. (1985). Summarising shapes numerically g-& h distribution. Exploring data Tables, Trends and Shapes:461-513
[7]. Shitan M. & Vazifedan T. (2012). Exploratory for almost anyone. 2nd edition, UPM Press
[8]. Siegel, A.F. (1983). Low Median and least Absolute Residual Analysis of Two-Way Tables. Journal of the American Statistical Association 78:382, 371-374.
[9]. Smith E.P. (2012). Baci design. Encyclopedia of environmentrics.
[10]. Sun, Y. & Genton, M. G. (2012). Functional median polish. Journal of Agricultural, Biological, and Environmental Statistics 17 354–376.
[11]. Tukey, J.W. (1977). Exploratory Data Analysis, Addison-Wesley, 231 32.
[12]. Velleman, P.E and Hoaglin,D.C (1981). Applications, Basics, and Computing of Exploratory Data Analysis. Duxbury Press