

Beckenham Place Park (III) | Lewisham Council | 2020  
The Lewisham Council

Beckenham Place Park (III)



## Data Analysis Tools User Guide

By:

John Trainor | AE | '21

## Table of Contents

Chapter 1. Purpose .....	1
Chapter 2. Background.....	2
2.1 Chi-Squared Test .....	2
2.2 Kruskal-Wallis Test .....	3
Chapter 3. Installation and Setup.....	4
Chapter 4. Using the Tools.....	4
4.1 Pivot Table Tool .....	5
4.2 Delimit Tool .....	7
4.3 Chi-Squared Test Macro .....	8
4.4 Kruskal-Wallis Test Macro .....	9

## List of Figures

Figure 1. Populating PivotTable Fields.....	6
Figure 2. Delimit Tool Typical Header Selection Order.....	8
Figure 3. Sample Interpretation of Chi-Squared Test Results.....	8
Figure 4. Typical Header Selection for Kruskal-Wallis Test .....	10
Figure 5. Sample Interpretation of Kruskal-Wallis Test Results.....	11

## List of Tables

Table 1. Factors of Interest .....	1
Table 2. Example of Contingency Table.....	3
Table 3. Data Analysis Preparations Checklist.....	5

## Chapter 1. Purpose

In a variety of industries and professions, statistical tests are used to examine large sets of data. The data collected by the Lewisham Council in Beckenham Place Park is no exception. Specifically, they are used to test a null hypothesis, which states that there are no significant differences between two or more sets of observed data. In the field of medicine, such tests are used to determine the effectiveness of experimental pharmaceuticals and treatments. In manufacturing industries, they can be used to test the impacts of new technology on output and quality. For the Lewisham Council, statistical tests will be used to determine how different visitors responded to the surveys distributed throughout the park, revealing any underlying trends. There are numerous factors of interest that may have an impact on responses to the survey questions. A number of these factors of interest are listed below in table 1.

*Table 1. Factors of Interest*

<b>Factors of Interest</b>
Age
Gender
Ethnicity
Disability
Residential Status
Frequency of visits to the park
Reason for visiting the park

To test for significance in the data with respect to these factors of interest, standardized statistical tests and a procedure are required. Having both will prevent inaccuracies in the results. For testing the survey and observational study data, Microsoft Excel will be used. Microsoft Office is a powerful tool for manipulating and analyzing data. Microsoft Excel allows users to

create custom tools and functions for niche tasks, such as statistical tests. Using this functionality, we have developed a series of tools to streamline and standardize the data analysis process for the Lewisham Council. These tools require very little prior experience with both Microsoft Excel and statistics. The following is a guide on how to implement and use these tools for future analysis.

## Chapter 2. Background

Within the Beckenham Place Park survey and observation sheets, there are several different types of responses. Among these are multiple-choice responses, check-all-that-apply (CATA) responses and ranking responses. The observational studies are completed by volunteers or members of the Lewisham Council. As such, each response on the sheet is completed with no blank responses left. The surveys, however, are completed by park visitors, who may leave one or more responses blank. Furthermore, visitors often differ from each other, whether it is by age, gender, or any number of additional categories. As a result, the sizes of data sets to be compared are often different from each other.

The Chi-squared and Kruskal-Wallis tests were chosen for their ability to analyze data sets of differing sizes. Additionally, each test can be used on multiple types of data sets, corresponding to each type of response found in the survey and observation sheet. Both tests indicate the presence of significant differences between the distribution of responses in each data set. A significant difference is a difference that is unlikely to be due to chance alone. The threshold for significance is dictated by Cronbach's alpha. For most social sciences, the value of alpha is set to 0.05, or 5%. Any probability less than or equal to 5% indicates a significant difference. Both statistical tests have limitations. The primary limitation of both tests is that they indicate the presence of significance, but do not pinpoint where the significance is. Therefore, they required additional tests, or post hocs, to determine which combinations have significant differences. The Chi-squared and Kruskal-Wallis tests use the adjusted residuals and Dunn's test respectively to do this.

### 2.1 Chi-Squared Test

The Chi-squared test can be applied to multiple-choice and check-all-that-apply responses. This test utilizes a contingency table, which shows the distribution of the frequencies of response. Each row and column in the contingency table represents a different category or

treatment. An example of a contingency table is shown below. The two inside rows represent two groups of patients 20 with the same illness; group A had pre-existing conditions and group B had no pre-existing conditions. The two inside columns represent the patients' responses to a new medication, which may be positive or negative. The numbers within the table represent how many patients of each group had each type of response to the medication, with the totals of the rows and columns on the outside.

*Table 2. Example of Contingency Table*

	<b>Positive Response</b>	<b>Negative Response</b>	<b>Grand Total</b>
<b>Group A</b>	1	9	10
<b>Group B</b>	7	3	10
<b>Grand Total</b>	8	12	20

In Microsoft Excel, this type of table is referred to as a pivot table. The Chi-squared test determines if there are significant differences between the observed response frequencies, such as those shown in Table 2, and the expected response frequencies. The test will not directly indicate which observed frequencies are significantly different from their corresponding expected frequencies. The adjusted residuals post hoc can do this. This post hoc adjusts the critical value below which the null hypothesis is rejected based on the number of combinations in the contingency table, which is four in the example Table 2, and compares it with the test statistics from the Chi-squared test.

## 2.2 Kruskal-Wallis Test

The Kruskal-Wallis test can be applied to ranked responses. The responses often follow a 1-10 scale or Likert scale (1-5). This test, which is an extension of the popular Mann-Whitney U test, determines if at least one data set is stochastically dominant over another. For ranked responses, stochastic dominance means that a data set has a significantly greater proportion of high ranked responses to low ranked responses than another set. The Kruskal-Wallis test will not indicate which pairs of data sets result in significance. A post hoc Dunn's test will determine

which pairs are significant. This test compares all possible combinations of data sets and determines if their respective test statistics are above or below the critical value.

### Chapter 3. Installation and Setup

The data analysis tools are packaged into a Microsoft Excel add-in. Installing the tools as an add-in will make them available in any future spreadsheet automatically. To install the add-in:

1. Open File Explorer and navigate to the **Documents** folder
2. In the **Documents** folder, create a new folder named **ExcelAddIns**
3. Download the file named **Statistics.xlam**, and move it to the **ExcelAddIns** folder
4. Open **Microsoft Excel**
5. Select **Options -> Add-ins**
6. Select **Go** at the bottom of the window, making sure that **Excel Add-ins** is select in the drop-down box
7. Select **Browse** and search for **Statistics.xlam** in the **ExcelAddIns** folder
8. Select **OK** and **Statistics** will appear in the list of available Add-ins with the box checked
9. Click **OK**





When a spreadsheet is opened in Excel after the add-in is installed, a tab named *Statistics* will appear next to the *Home* tab. In this tab are all the necessary tools for running the statistical tests, except for the pivot table tool.

### Chapter 4. Using the Tools

Before running a statistical test, the data must be formatted and prepared properly. The figure below is a checklist that you should run through before analyzing the data. The checklist includes renaming the data sheet, formatting the data into a table, and adjusting the widths of each column. Doing this will not only allow the tools to work more effectively but will make the data easier to read and sort through as well.



Table 3. Data Analysis Preparations Checklist

Action Item	Instructions
<p>1. Rename the sheet containing the data (This creates a safeguard to prevent the data from accidentally being deleted or overwritten in the event of an accident)</p>	<p>1. In the bottom left-hand corner, right-click on the sheet containing the data and select <b>Rename</b></p> <p>2. Type “<b>Responses</b>” and press <b>Enter</b></p>
<p>3. Format data into a table it is not already</p>	<p>1. Select cell <b>A1</b> and press <b>Ctrl+A</b> to select all the data</p> <p style="text-align: right;">  </p> <p>2. In the <b>Home</b> tab, click on  and select a template</p> <p>3. Make sure the box next to <b>My table has headers</b> is checked</p> <p>4. Press <b>Yes</b> if any additional messages pop up asking if you want to continue</p>
<p>4. Auto-fit column widths</p>	<p>1. Select cell <b>A1</b> and press <b>Ctrl+A</b> to select all the data</p> <p style="text-align: right;">  </p> <p>2. In the <b>Home</b> tab, click on  and select <b>AutoFit Column Width</b></p>

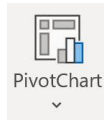
#### 4.1 Pivot Table Tool

Unlike the other tools, the pivot table tool already exists in Microsoft Excel. This tool creates a contingency table out of the data in another table, which in our case would be the table of survey or observational data. To create a pivot table:

1. Select a cell within the table you want to analyze



2. In the **Insert** tab, select



3. Alternatively, click the drop-down on **PivotChart & PivotTable**, which will include an interactive graph.

This will create a pivot table, and graph if chosen, in a new sheet. At the same time, a field list will appear on the right side of the screen. This list will appear every time you click on the pivot table. If it does not appear when you click on the pivot table, select **Field List** in the **PivotTable Analyze** tab. The fields should be the headers of the table you are analyzing. To add fields to the pivot table, simply drag and drop them into the areas below the list of fields. The **factor of interest**, such as age or gender, should be dragged and dropped into the **Rows** area. Likewise, the question should be dragged and dropped into both the **Columns** and **Values** areas. This is demonstrated in the figure 1, below.

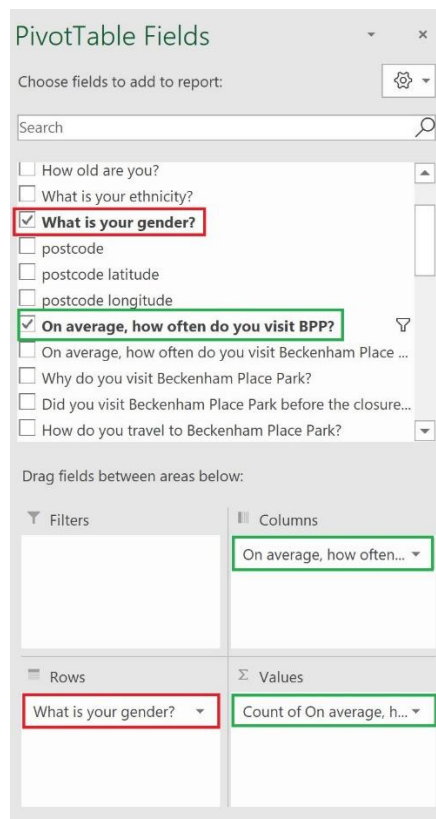


Figure 1. Populating PivotTable Fields


For each field that is selected, it is important to filter out blank responses if there are any, as they will skew the data. They can be filtered using the dropdowns next to **Column Labels** and **Row Labels** in the pivot table itself. Blanks may say (*blank*) or will have no text at all. To add different fields to the pivot table, simply uncheck selected fields in the fields list, and drag and drop new ones.

#### 4.2 Delimit Tool

Check-all-that-apply responses cannot be immediately added to a pivot table, as each cell often includes two or more responses separated by commas. Each cell must be separated so that there is one response per cell instead. The delimit tool can be used to do this automatically. To use this tool:

1. Select the **Responses** sheet



2. Click on  in the **Statistics** tab
3. Select the header of a **factor of interest** (see figure 2 for reference) and click **OK**
4. Select the header of a **check-all-that-apply question** (see figure 2 for reference) and click **OK**

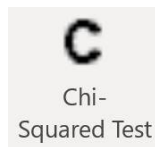
The order in which the headers are selected is crucial to the function of the tool. Figure 2 shows an example of the typical order in which the headers should be selected. If a user wishes to compare what visitors do in the park based on their location in the park, the header of the column with the locations in the park must be selected first, and then the header of the column indicating what the visitors do must be selected last, as it contains multiple responses per visitor. After the headers are selected, a new sheet will be created with a table containing the delimited data. This table can then be selected to create a pivot table.

L		M	
1	Location in the Park	2	Why do you visit Beckenham Place Park?
Zone 5 (West fields)			Dog Walking,Relaxation
Zone 3 (Woods)			Dog Walking,Walking,Relaxation,Spending time outdoors
Zone 4 (Mansion)			Dog Walking,Walking
Zone 4 (Mansion)			Dog Walking
Zone 4 (Mansion)			Walking,Dog Walking
Zone 5 (West fields)			Walking,Cycling
Zone 4 (Mansion)			Dog Walking
Zone 3 (Woods)			Walking
Zone 4 (Mansion)			Dog Walking,Walking
Zone 4 (Mansion)			Dog Walking
Zone 4 (Mansion)			Daughter does art classes in mansion
Zone 2 (East fields)			Dog Walking,Walking,Wildlife,Relaxation,Spending time outdoors,Other sport/ exercise,Running

Figure 2. Delimit Tool Typical Header Selection Order

### 4.3 Chi-Squared Test Macro

This macro can be used to automatically perform a Chi-squared test on the data in a pivot table. This macro assumes that the user has already created a pivot table, added their desired fields and filtered out blank responses. To run the macro:



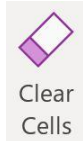
1. Select **Chi-Squared Test** in the **Statistics** tab

The macro will automatically calculate the expected frequencies and determine if there is significance. A message box will appear automatically if no significance was detected. If significance was detected, the post hoc will run automatically. Areas where there was significance will be highlighted in red or green. Cells highlighted in red indicate areas where the number of observed responses was far greater than expected. Likewise, cells highlighted in green indicate areas where the number of observed responses was far less than expected. This is demonstrated in figure 3, below. The cell highlighted red indicates that more females responded “None” than was expected. The cell highlighted in green indicated that less males responded “None” than was expected.

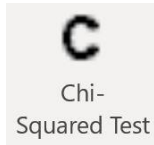
Adjusted Residuals	1/year	2/year	3-4/year	5-6/year	7-12/year	None
Female	-1.920139366	-1.034501912	-0.121706107	0.961121046	-2.128196603	2.818175795
Male	1.920139366	1.034501912	0.121706107	-0.961121046	2.128196603	-2.818175795

Figure 3. Sample Interpretation of Chi-Squared Test Results

To run a subsequent Chi-squared test on different data:



1. Select **Clear Cells** in the **Statistics** tab
2. Replace the fields in the pivot table with new ones
3. Filter out any blank responses

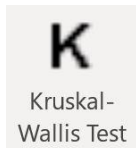


4. Select **Chi-Squared Test** again

#### 4.4 Kruskal-Wallis Test Macro

This macro can be used to automatically perform a Kruskal-Wallis test on data directly from the table in the **Responses** sheet. To run this macro:

1. Select the **Responses** sheet



2. Select **Kruskal-Wallis Test** in the **Statistics** tab
3. Select the header of the **factor of interest** (see figure 3 for reference)
4. Select the header of **ranking question** (see figure 3 for reference)

The order in which the headers are selected is critical for the tool to execute the test properly. Figure 4 shows provides an example of the order in which the headers should be selected. If a user wishes to compare how visitors feel about the park based on how often they visit, the header of the column indicating how often they visit would be selected first, and the header of the column indicating how they feel about the park would be selected last.

1	Q	R
	On average, how often do you visit BPP?2	2 How do you feel about the park as it is now? (Ranked)
Everyday		9
Once or twice a week		8
Everyday		8
Once or twice a week		8
Once a month		6
Once or twice a week		8
Once a month		5
Once every six months		8
Everyday		2
Once a month		8
Once a month		9
Once or twice a week		7
Once or twice a week		3
Once a month		8
Once or twice a week		8
Once a month		8
Everyday		4
Once or twice a week		10
Once or twice a week		8
Once or twice a week		9
Everyday		8
Once a month		10
Everyday		10
Everyday		10
Once or twice a week		8
Everyday		9
Everyday		7
Less than once a year		7
Everyday		7
Once or twice a week		8

Figure 4. Typical Header Selection for Kruskal-Wallis Test

After the proper headers are selected, the macro will automatically create a new sheet and perform the calculations. A message box will appear automatically if no significance has been detected. If significance is detected, a new sheet will be created automatically with the post hoc. In this sheet, the results will be tabulated and any combinations with significant differences will be highlighted green. An example of this is shown below in figure 5. The results shown in figure 5 are from a test comparing understanding of the heritage of Beckenham Place Park between visitors of different ethnicities. The results show that there are significant differences between “**White British**” and “**Black/African/Caribbean/Black British**” visitors, as well as between “**White Other**” and “**White British**” visitors.

	Asian/Asian British	Black/African/Caribbean/Black British	Other	White British	White Other
Asian/Asian British	-				
Black/African/Caribbean/Black British	No	-			
Other	No	No	-		
White British	No	Yes	No	-	
White Other	No	No	No	Yes	-

*Figure 5. Sample Interpretation of Kruskal-Wallis Test Results*

Before running a subsequent Kruskal-Wallis test on different data, be sure to:

1. Delete the two sheets from the previous test using
2. Select the **Responses** sheet again

