Introduction

In this document, we will describe several more advanced features of Excel. In our opinion, these are especially useful when one is faced with analysis of large datasets, such as might be encountered when analyzing EMS data. Again, there are many other programs and utilities that are used to analyze data. We have chosen to illustrate EMSC data analysis using Excel because the program is widely available, fairly intuitive to use, and has capabilities that are useful to EMS data analysis.

Educational Objectives

After reading this document, you will:

1. Understand how to use Excel as a "front end" to databases contained within other programs such as SQL Server, Oracle, or Access.
2. Appreciate the utility of PivotTables, and be able to set one up.
3. Know how to invoke the data analysis tools available to run statistics within Microsoft Excel.
Connecting to External Databases

In our opinion, Excel is a superb tool to use as a "front end" to large databases. To accomplish this, you must have an ODBC connection set up. This will not be demonstrated in this workbook because setting up the ODBC requires specific knowledge of databases, and requires setting up computer permissions to allow the user access to the database.

The main advantage of connecting directly to the databases is to identify only the data you want, and then have it sent you. This is in contrast to asking the database server to send you everything and let you sort it out in Excel. Suppose you have a million EMS records. Your spreadsheet can't even hold that many rows, and you are really only interested in 5 year old victims of drowning. By using Excel properly, you can request 5 year old drowning victims, the database will do the work of finding them, and then only those records will be sent to your Excel spreadsheet.

Once an ODBC is set up, you can connect to an external database. This is accomplished by first clicking on “Data” – “Get External Data” – “New Data Query” (Figure 1). The next step will be to choose the data source you wish to access. For this example, we will be showing how to retrieve a Microsoft Access Database. However, the steps for accessing other database managers are not much different and Excel walks you through nicely. To retrieve data from an Access Database, you need to click on the “MS Access Database” in the Choose Data Source box. It will then ask you to specify the file location of the Access Database you wish to use.

Figure 1: Bringing external data into Excel
After you have connected to the data source, Microsoft Query will begin. Since Microsoft Query is also used in PivotTables and Charts, we will spend some time discussing this product. The first box specifies the columns you wish to include in your query (Figure 2). In this box you are also able to preview the data elements in each field. Clicking “Next” will then bring up the box for filtering the data (Figure 3). For our example, we restricted the data to those records where the age is between 0 and 17 years.

The next box allows you to sort the data by any column. The final box has several different options for handling the data. The default is to return it to Excel, but you are also allowed to choose to further edit the data in Microsoft Query. If you chose to edit it further, you will be able to write your own SQL script. Figure 4 contains a screen shot of the returned data in Excel.

<table>
<thead>
<tr>
<th>A</th>
<th>B</th>
<th>C</th>
<th>D</th>
<th>E</th>
<th>F</th>
<th>G</th>
<th>H</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Date</td>
<td>IDNum</td>
<td>InsDate</td>
<td>Age</td>
<td>County</td>
<td>Gender</td>
<td>Event</td>
</tr>
<tr>
<td>2</td>
<td>97</td>
<td>9716</td>
<td>7/19/1997</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>Vital Assessed-Monitored</td>
</tr>
<tr>
<td>3</td>
<td>97</td>
<td>9716</td>
<td>7/19/1997</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>Vital Assessed-Monitored</td>
</tr>
<tr>
<td>4</td>
<td>97</td>
<td>9716</td>
<td>7/19/1997</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>Vital Assessed-Monitored</td>
</tr>
<tr>
<td>5</td>
<td>97</td>
<td>9716</td>
<td>7/19/1997</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>Vital Assessed-Monitored</td>
</tr>
<tr>
<td>6</td>
<td>97</td>
<td>9716</td>
<td>7/19/1997</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>Vital Assessed-Monitored</td>
</tr>
<tr>
<td>7</td>
<td>97</td>
<td>9716</td>
<td>7/19/1997</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>Vital Assessed-Monitored</td>
</tr>
<tr>
<td>8</td>
<td>97</td>
<td>9716</td>
<td>7/19/1997</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>Vital Assessed-Monitored</td>
</tr>
<tr>
<td>9</td>
<td>97</td>
<td>9716</td>
<td>7/19/1997</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>Vital Assessed-Monitored</td>
</tr>
</tbody>
</table>

Figure 2: Choosing columns in Microsoft Query

Figure 3: Filtering the data in Microsoft Query

Figure 4: The Excel table that was generated by retrieving data from Access.
PivotTables

Filters and formulas are great, but PivotTables offer an alternative that is often more useful. PivotTables are an excellent way to quickly summarize large amounts of data and display them in relative terms in a table format; PivotTable are very good for the kind of data analysis that EMS workers often need to do. Creating a PivotTable is a bit less intuitive than working directly on the spreadsheet with filters and formulas; data analysis with PivotTables creates a new table entirely, that often appears on a different worksheet than the data analyzed.

Using PivotTables requires that your worksheet be properly set up with some entry in every cell in the top row of the data table to be analyzed (this entry will serve as the variable name in the PivotTable). To start the PivotTable platform click on “Data” → “PivotTable and PivotChart Report” (just PivotTable Report in some versions). The first dialog box allows you to choose how you will get the data and whether you want a pivot table or a pivot chart. In addition, when you use Excel to access large external databases, PivotTables are often much more practical than trying to read the database values into a huge spreadsheet. The options for where the data can be obtained include an Excel table and an external data source. If you choose an external data source you follow the same steps as mentioned above in connecting to an external data source.

Exercise: Use a PivotTable to analyze the proportion of EMS runs done by paramedics and EMTs in the various counties. (If you start by selecting the entire data table, Excel will pre-fill the data source box for you. Else, or if you are using an external data source, you can specify it later.)

- Select Data Menu  ➔  PivotTable Report and PivotChart Report
• The following dialog box comes up on the data table:

• Choose “Microsoft Excel database” for the data source and “PivotTable” for the report that you want to create. Then click “Next”.

• The Step 2 dialog box appears.
• The data source range may already be filled in, as it is here, or you can fill it in. This can be done by clicking & dragging in the data table if you like; you can collapse the dialog box to get it out of the way by clicking in the little box to the right of the data range entry box.

• Click Next, and the step 3 dialog box appears. Here, just click on “New worksheet”, click “Finish”, and the PivotTable appears on a new, blank worksheet.
• The PivotTable is also blank because you haven’t yet chosen the data that you would like it to analyze. This is done by dragging items from the Field List onto the PivotTable area.
• Note that the word “field” is used for variables; means the same thing. (Field or variable names are the headers for the data columns on the main spreadsheet.)
• Here, we are planning to analyze the number of runs of various characteristics, so you can choose any variable (column) that has an entry for every record.
• Click on Pt_num (which is just an identifier and so exists for every record) and drag it into the “Drop Data Items Here” area of the PivotTable.
• You haven’t yet told Excel how you would like to express these data, so it chooses “Sum of EMS num” for you. This is not helpful, since you want to analyze counts. Right click or double click on the cell containing “Sum of EMS num”.

```
<table>
<thead>
<tr>
<th></th>
<th>A</th>
<th>B</th>
<th>C</th>
<th>D</th>
<th>E</th>
<th>F</th>
<th>G</th>
<th>H</th>
<th>I</th>
<th>J</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td></td>
<td></td>
<td>Drop Page Fields Here</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td></td>
<td></td>
<td>Drop Page Fields Here</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4</td>
<td></td>
<td></td>
<td>Drop Row Fields Here</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5</td>
<td></td>
<td></td>
<td>Drop Column Fields Here</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>6</td>
<td></td>
<td></td>
<td>Drop Data Items Here</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>19</td>
<td></td>
<td></td>
<td>PivotTable</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>20</td>
<td></td>
<td></td>
<td>PivotTable</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>21</td>
<td></td>
<td></td>
<td>PivotTable</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>22</td>
<td></td>
<td></td>
<td>PivotTable</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
```

![Excel screenshot with PivotTable and Field List](image)
This will enable you to bring up the PivotTable Field dialog box, that will let you choose different ways of summarizing data.

- Drag down to “Count” and click OK.
- “Sum of EMS num” will change to “Count of EMS num”.
- Note that you can also rename the field in this dialog box by typing in the “Name” box.

- From the PivotTable Field List, drag “age” over to the column part of the PivotTable (the cell directly under “Count of Pt_Num”).
- The PivotTable then fills out and shows you the number of runs for each age contained in the data, as shown here.
Exercise: Try analyzing a different variable and expressing it as a percent of the total.

- Click on the “age” cell in the PivotTable and drag it back to the Field List.
- Drag “Prov_Imprs” to the column area of the PivotTable.
- Then, right click on the cell containing “Count of Pt_Num” and bring up the Field Settings dialog box.
- Click on “Options”.
- Click on “% of total” and click OK.

Result: Values are now expressed as percent of total (right).

Important: When creating a new PivotTable, use an existing PivotTable as the data source if possible (it’s a choice in the PivotTable Wizard). Your file will become excessively large otherwise.

Exercise: Try making a table with two data columns.

- Create a PivotTable and drag “Pt_Num” into the data area twice.
- You should see a table as shown here.
- Click on the cell containing the word “Data” and drag it one cell rightward.
- The two “Pt_Num” data rows have become columns.
- Use the Field Setting dialog box to change one column to Count and the second column to % of total.
- Drag “Prov_Imprs” into the row area.
**Result:** The final result should resemble the table on the showing two data columns.

Note: It may be difficult to undo the two data column arrangement.

**Exercise:** Create a table and eliminate blank cells from analysis.

- Create a table analyzing the variable “med” as below left. Click on the little arrow on the right hand side of the cell containing the word “med”.
- A list comes up of variables to show. De-select for “blank”.
- You can also right-click in the cell you want to hide.

**Result:** A table like the one on the right below. This may be useful if you want to remove blank or null values from the denominator.
Exercise: Create a two-variable PivotTable.

- Create a PivotTable for Prov_Imps
- Drag “gender” into the column area.
- This should yield a table like the one on the right.

Exercise: Try adding a second variable to the column area.

- Drag “med” into the column area.
- The result is shown below. It is messy, too big to see at once.
- There is an alternative way to show additional detail.
**Exercise:** The settings of the row and column fields can be changed by right clicking and then clicking on “Field Setting.” For example, you can change the pivot table to show only the top 10 injury/illness codes in terms of frequency.

- Right-click the Prov_Imprs cell.
- Click on “Field Setting”.
- Click on “Advanced” in the field setting dialog box.
- The “Advanced Options” dialog box appears.
- To show only the top 10, click on “Automatic” under “Autoshow Options”.
- Click OK and the PivotTable shown below on the left appears, showing only the top ten.

To sort these data in order of most frequent:
- Go to the “Advanced Options” dialog box as noted above.
- Click on “Descending” under “Autosort”. The PivotTable becomes the one on the right below:
Drag “med” back to the Field list and doubleclick on a row item in the PivotTable.

- Select any of the variables to get more detail.

**Result:** Here is shown the result of expanding detail for the variable “med” for patients with Prov_Imprs of “Traumatic injury”.

---

**Image Description:**

- The image shows a PivotTable displayed in a software application, possibly Microsoft Excel, with data related to traumatic injury and various medical interventions.

- The table includes columns for variables such as Prov_Imprs, General-Medications, and other medical interventions.

- The PivotTable is showing detailed data for a specific patient record, with fields like medication type, count, and details.

- The interface includes options for expanding and collapsing details, with pop-up windows for selecting and showing additional data.

---
Exercise: Grouping. Sometimes you may want to group several items into one. For instance, the PivotTable at right shows several Prov-Imprs entries that are related to head and neck problems. You might want to make a new item that summarizes all 3.

- Select the categories you want to group.
- Right-click on the selection to bring up a menu box.
- Select Group and Outline → Group

Result: A new item is created that is a grouping of the 6 selected, as seen below. Excel names it “Group 1”, it can be renamed in the formula bar. It also appears as a new item in the Field List.

The detail can be collapsed by double-clicking on “Group 1”, to yield the result seen at the bottom of the page.
We can also change the values types displayed in the data cells. For example, we can have the data values be the average age. This is done by first changing the data field to be age then choosing average under the field setting.

**Table of average age by diagnosis and gender.** Here, age is the data field, gender the column field, inj_code the row field. Average is selected in the Field Setting dialog box.

Another nice feature in PivotTables is the page field, a spot at the top of the PivotTable. This allows you to put another data element in as a “Page”. Adding a page value allows you to filter the table by this value. In this example, we have put Incid_Cnty into the Page Field enabling us to display only records from an individual county. You can also right click on the table and choose “Show Pages” to create this table in separate worksheets, one for each value of the page field.
Other things about Excel PivotTables:

- PivotTables can be named. Right click anywhere in the table and bring up the “Table Options dialog box.
- Data are still connected to the database. This allows you to update the tables when the database changes.
- A potentially useful feature is called “drilldown”.
- Try double-clicking on any cell in a PivotTable.
- A new spreadsheet will appear populated only by the records referred to in that cell. You might want to do this if you want to concentrate on a certain subset of the data that you have identified via the PivotTable analysis.
Microsoft Excel Data Analysis Add-In

The normal installation of Excel does not automatically load the Data Analysis Add-In. To see if your machine has this loaded, check the “Tools” menu to see if it is available. This is shown below—note that “Data Analysis” is listed as the last tool.

If you do not see this item then you must install the add-in. It is already on your hard drive—you merely have to do a few simple things. Select the “Tools” – “Add-Ins” – “Analysis Toolpak”. While you are there, make sure the ODBC Add-In is also selected so you can connect to an external database. Click OK and you're done.
Now that the add-in is available, click “Tool” - “Data Analysis” and you will get a data analysis dialog box.

This dialog allows you to access a large variety of common statistical procedures. Each works with special wizards to help you figure out how to set up the analysis. If you selected descriptive statistics, for instance, you would see the wizard that is shown below.

If, of course, you aren’t sure what you are seeing, click on the “Help” button and this ties you into the extensive on line documentation of the program. If you still need help, please feel free to contact NEDARC at (801) 585-9761 or www.nedarc.org for further assistance.