

# Inference in Excel QuickStart 4

## *Use Inference Table DataFrames in Inference in Excel*

Inference in Excel QuickStart 4 shows you how to:

- Create a DataFrame within an Inference in Excel document.
- Specify data vectors and add data.
- Specify the data type and additional attributes for data vectors.

### What is a DataFrame?

Inference DataFrames QuickStart 1 covers the basics of DataFrames, which are rectangular data structures similar to spreadsheets, database tables, matrices and rectangular lists.

Here is a sample Data Frame:

	Temp	Conc	Cat	Yield
1	160	20	A	60.1
2	180	20	A	72.5
3	160	40	A	54.2
4	180	40	A	68.7
5	160	20	B	52.4
6	180	20	B	83.1
7	160	40	B	45.7
8	180	40	B	80.3

In Inference, DataFrames are represented as software objects with an API (application programming interface) that allows you to programmatically analyze and manipulate their content.

### How do I create a DataFrame?

Inference provides two ways to create a DataFrame:

- **As a separate document**  
Inference DataFrames QuickStart 1 shows how to add an Inference DataFrame container to an Excel document.
- **Directly in an Inference in Excel parts container document as an Excel Table/List**  
Excel 2007 allows you to turn a range of cells into a Microsoft Excel table. In Excel 2003, this feature is referred to as an Excel list. When a table (or list) is created on the Input worksheet of an Inference in Excel document, Inference automatically parses the data into R code stored as an object named **ExcelInputs**. R can use this object to instantiate an R DataFrame object.

Using this approach provides you both the advantages of Excel tables/lists (managing the data in the table rows and columns independently from the data in other rows and columns on the

worksheet) and the advantages of ad-hoc R DataFrames created on the fly for use in R analyses.

**Note:** You can view descriptions of Excel tables at <http://office.microsoft.com/en-us/excel/CH101024591033.aspx> or view a short video demonstrating the use of Excel tables at <http://office.microsoft.com/en-us/excel/HA102376171033.aspx>.

## Which approach should I use?

Use the Inference in Excel DataFrame Container approach when you:

- Need to create and use DataFrames in both Inference in Word and Inference in Excel documents
- Need to create DataFrames from SDML files written by third-party devices or systems (e.g., MATLAB) or custom applications using the SDML object API; or
- Need the extended properties and attributes capabilities of Inference DataFrames only available through the SDML file.

Use the Inference Excel table/list approach when you:

- Need a rapid, single-step means to assemble a DataFrame for use in Inference Excel only; and
- Want to use the full capabilities of Excel tables/list to manage the content of the DataFrame.

## Creating an Excel Input Table/List and Mapping It to a DataFrame

### 1. Add a Parts Container to an Excel Document

An Inference in Excel document is a standard Excel document with an embedded Parts Container. To add a Parts Container to an Excel document:

1. Open or create a new Excel document.
2. In Excel 2007: Click the **Inference** tab on the Excel Ribbon. Click **Add Container**, then select **Add Parts Container**.  
In Excel 2003: On the **Inference** menu, select **Add Parts Container**.
3. The **Inference in Excel Task Pane** and two new worksheets (labeled **Input** and **Output**) will appear.

### 2. Place Data on the Input Sheet

Data on the Input sheet should be organized in columns, with a label in the first row of each column. For this example:

1. Select and copy the four columns (Temp, Conc, Cat, Yield) from the sample data set above.
2. On the Task Pane of the Inference in Excel document, click to open the **Inputs** section.
3. On the **Inputs** worksheet, paste the sample data set to cell A1. Compare your results to the following:

	A	B	C	D	E	F	G	H	I
6		Temp	Conc	Cat	Yield				
7	1	160	20	A	60.1				
8	2	180	20	A	72.5				
9	3	160	40	A	54.2				
10	4	180	40	A	68.7				
11	5	160	20	B	52.4				
12	6	180	20	B	83.1				
13	7	160	40	B	45.7				
14	8	180	40	B	80.3				
15									
16									
17									
18									
19									
20									
21									
22									
23									
24									
25									
26									
27									

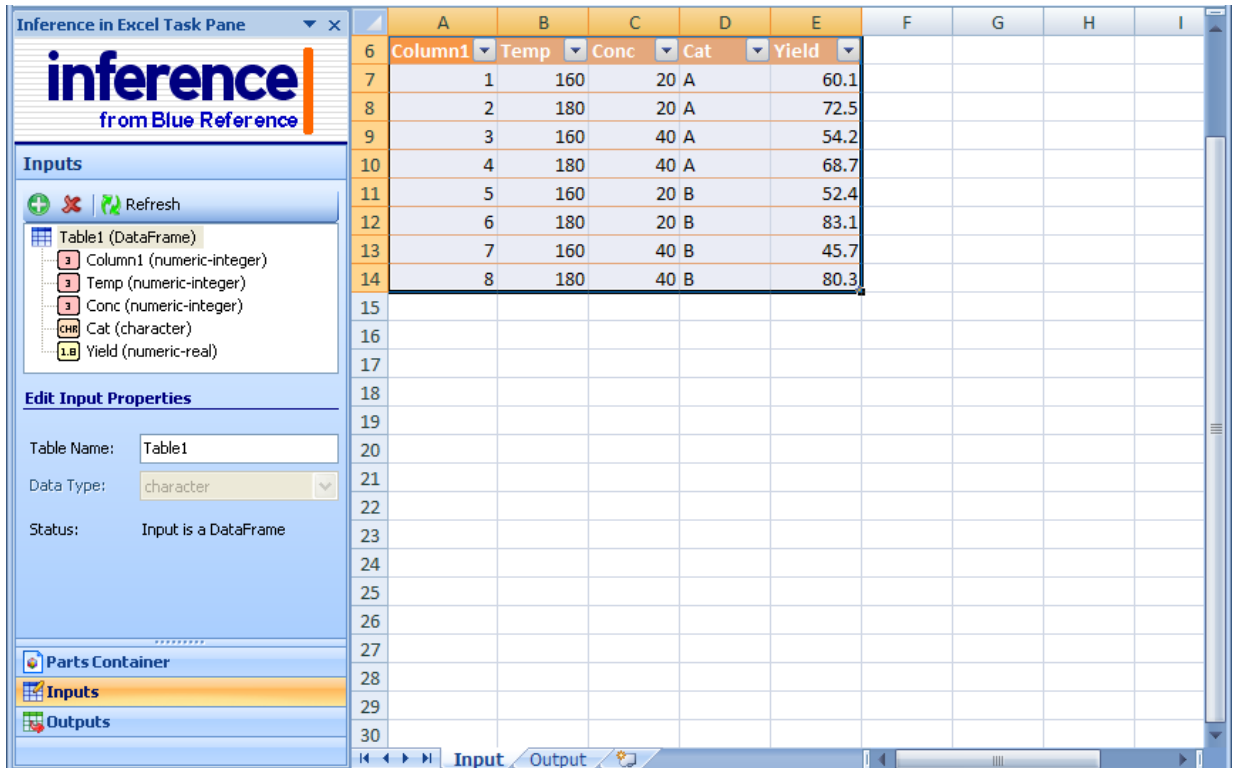
### 3. Create an Excel Table/List

To create an Excel Table/List from the data you just added to the Input worksheet:

1. Select the **Input** worksheet.
2. Select the range A1 to E9 covering the pasted data
3. In Excel 2003: Press **Ctrl+L** to create a list. In the Create List dialog, enable the **My list has headers** option.

In Excel 2007: Select **Home > Styles > Format as Table** and pick the desired table style. In the next dialog box, confirm that the **My table has headers** is checked. Click OK.

4. On the **Input** section of the Inference Task Pane, click **Refresh**.
5. Compare your results to the following:



#### 4. Edit Input Properties of the Mapped DataFrame

When Inference generates the inputs from the table (list), it automatically assigns a table name, column names, and data types. These properties are easy to change.

**Table (DataFrame) Name.** Inference assigns a default range name to the table and the corresponding DataFrame (Table1, Table2, ...). To change the name of the table from the default, **DataFrame**, to **MyDataFrame**:

1. Select the **Input** section of the Task Pane.
2. On the Inputs list, select the name of the table (DataFrame) you want to edit.
3. In the **Input Properties** section of the Task Pane, click Edit.
4. Enter **MyDataFrame**, then click **OK**.
5. To ensure that the Task Pane reflects the current status, click on the **Refresh** button on the **Inputs** menu.

**Column (Vector) Name.** Columns are assigned default names (Column1, Column2, ...) unless the range has specified headers. You can change the name of a column (vector). For example, to change the name of the column "cat" to "catalyst":

1. Select the **Input** section of the Task Pane.
2. On the Inputs list, select the name of the column (Vector) you want to edit.

3. In the **Input Properties** section of the Task Pane, click Edit.
4. Enter the new column name, then click **OK**.
5. To ensure that the task pane reflects the current status, click the **Refresh** button on the **Inputs** menu.

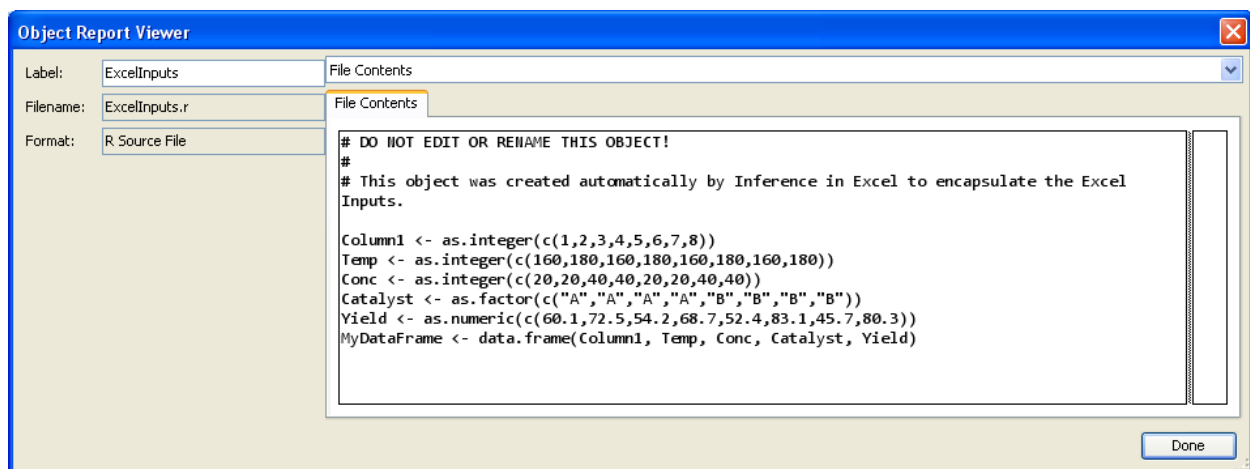
Alternatively, you can select any table header in the **Input** sheet and edit its value directly in the Excel Formula Bar. Again, to ensure that the task pane reflects the current status, click on the **Refresh** button on the **Inputs** menu.

**Column (Vector) Data Type.** When mapping a range in the Input sheet to a DataFrame, Inference makes inferences about the data in the table and assigns data types. Although the assigned data types are valid, they may not be the intended data type. For example, the column named “Cat” (catalyst) is assigned the character data type, but the intended data type is categorical. To set the column vector to the desired categorical data type:

1. Select the **Input** section of the Task Pane.
2. Select the name of the column (Vector) you want to modify – for this example, select **Catalyst**.
3. Under **Input Properties**, change the **Data Type** from **character** to **categorical**.
4. Click the **Refresh** button on the **Inputs** menu.

## 5. Use R to Instantiate and Access the DataFrame Object

Inference uses an object in the Parts Container to store the contents and associated information about the mapped DataFrame. This object is named **ExcellInputs**. The object is comprised of an R code file, which is updated every time the Inference in Excel document is saved. To view the contents of **ExcellInputs**, right-click and select **Open in Other Interviewer > Open in Object Report Viewer**.



NOTE: Do not remove or edit the content of **ExcellInputs**; this will yield unexpected results.

Using the R code in **ExcelInputs**, Inference will instantiate each DataFrame and each Vector declared in the **Inputs** section during the **Execute Document** process. To illustrate this capability, do the following:

1. Select the **Parts Container** section of the Task Pane
2. Right-click **Code Blocks** then select **Insert and Edit New Code Block**.
3. In the **Code Block Editor** add the following R instructions:

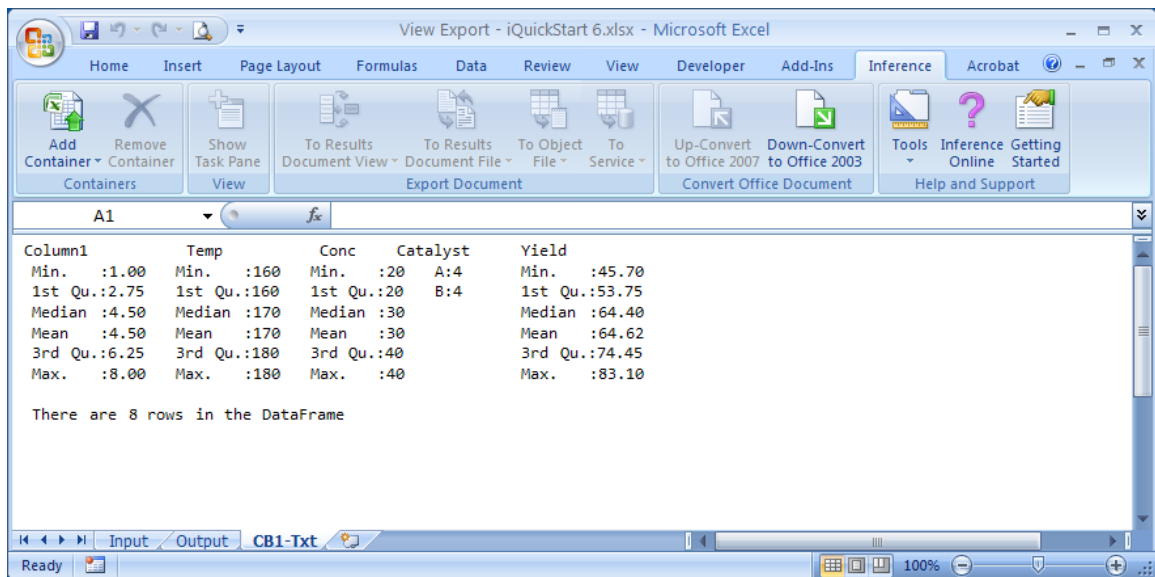
```
summary(MyDataFrame)
cat("\n", "There are", length(Column1), "rows in the DataFrame", "\n")
```

4. To view the results of executing this code, execute the document.

In Word 2007: Click the **Inference** tab on the Word Ribbon, then select **To Results Document View > Microsoft Excel View**.

In Word 2003: Click the **Execute Document** button at the bottom of the Inference in Word Task Pane, then select **To Results Document View > Microsoft Excel View**.

5. Compare your results to the following:



## 6. Manage the Data in the DataFrame

Excel is able to determine where the table/list ends. Thus, once the Excel table/list is mapped to the DataFrame, modifications to the Excel table/list will automatically be mapped to the DataFrame. This capability of Excel allows you to easily manage the data in a DataFrame in the following ways:

**NOTE:** The actions described below change the Excel Table/List and the mapped DataFrame assembled in **ExcelInputs**. To ensure that the contents of **ExcelInputs** reflect the current state of the Excel Table/List, either perform a **Save** on the file menu or a **Refresh** on the Inputs section of the Task Pane

before inspecting the **ExcellInputs**.

- **Insert Rows and Columns into the DataFrame.** Select the cell corresponding to the first row in the Yield column (E2). Go to **Home > Cells > Insert**. You can choose from **Insert Table Rows Above**, **Insert Table Column to the Left**, or **Insert Table Column to the Right**. Execution of the insert(s) will automatically add the rows and columns to the DataFrame and fill the empty spaces with NAs.
- **Delete Rows and Columns from the DataFrame.** Select the cell corresponding to the first row in the Cat (catalyst) column (D2). Go to **Home > Cells > Insert**. You can choose from **Delete Table Rows** or **Delete Table Columns**. Execution of the deletion(s) will automatically remove the selected row(s) or column(s) from the DataFrame.
- **Expand the DataFrame.** Placing the cursor at the bottom-right corner of the Excel Table/List changes the cursor to a double-headed arrow, which you can use to expand the DataFrame with additional empty columns and rows.
- **Remove Duplicate Rows from the DataFrame.** Select any cell in the Excel Table/List. Go to **Table Tools > Design > Tools > Remove Duplicates**. Check all the column boxes by clicking on **Select All** and click on **OK**. This action removes the duplicate rows from the Excel Table/List and the corresponding DataFrame. If you change your mind, you can back out (undo) removing the duplicates. If you wanted to first take a look at the duplicate rows, you can use the “Highlight Duplicate Values” feature that is part of conditional formatting.
- **Add a Calculated Column to the DataFrame.** You can quickly create a calculated column in the Excel Table/List and corresponding mapped DataFrame. A calculated column uses a single formula that adjusts for each row. It automatically expands to include additional rows so that the formula is immediately extended to those rows. You need to enter the formula only once; you do not need to use the Fill or Copy command.
- **Sort Data in DataFrame.** You can easily sort the data in the DataFrame two ways. One, you can select the desired column sort order from the dropdown boxes in the Excel Table/List header. Or two, you can select a cell in the Excel Table/List and go to **Home > Editing > Sort & Filter > Custom Sort** and specify the desired sort actions in the dialog box.

## 7. Control Display of Data in the Excel Table/List

Excel Tables/Lists provide an easy means to only display rows of interest. You can filter on multiple criteria or on icon sets applied via conditional formatting. It is important to recognize that the filtering action does NOT affect the underlying data in the data frame. It only affects the display of the data in the Excel Table/List—that is, the contents of **ExcellInputs** remain unchanged. You can filter display of data by selecting the desired filter action from the dropdown box in the Excel Table/List header.

Excel Tables (2007) also come with a large library of styles that you can use to apply attractive and consistent formatting to Excel Tables. You can apply a style by selecting a cell in the Excel Table/List and going to **Table Tools > Design > Table Styles > Quick Styles** on choosing from the range of available styles.

## 8. Convert an Excel Table/List to a Range of Data

At times, you may want to convert an Excel Table/List back to a standard range of data—that is, you want to unlink the **Input** data from the corresponding DataFrame assembled in **ExcelInput**.

1. Click anywhere in the Table. This triggers display of **Table Tools**, adding the **Designs** tab.
2. On the **Design** tab, in the **Table Tools** group, click **Convert to Range**.
3. You can also right-click the table, point to **Table Tools**, and click **Convert to Range**.
4. Note that Table features are no longer available after you convert the table back to a range.