

## The Power of Excel 2007: Identifying Duplicates



For years I searched for a reliable way to identify duplicates in an Excel spreadsheet. I tried everything I knew, including MS Access's "Find Duplicates Query Wizard" - to no avail. Finally, through networking at a conference, I came upon the answer below. Finding duplicates in Excel a three step process: inserting a new column with a formula, and then (here's the key), turn Automatic Calculation OFF and Manual Calculation ON so that when you Sort on the new duplicates column, the formula does not recalculate automatically. Lastly, once you've identified the duplicates and deleted them (or copy/pasted the unique's), you delete the column with the formula and turn Automatic Calculation back on.

Here's what to do:

- **Sort the datasheet by the field that has possible duplicates** (EmplID, SSN, for example) and then **Insert a column** to the left of the field that has the possible duplicates.
- In the 1<sup>st</sup> cell in the new column **type a heading** (Dups) adjacent to the duplicated field (SSN) field header. Then **type the following formula**: =IF(B2=B1,"DUP","") in the cell adjacent to the 1<sup>st</sup> possible duplicate, substituting your column name (A, B, C) in the formula. **Note**: the last part of the formula after the coma is double quotes ("").
- **Copy the formula** down to the appropriate cells – **Here's a cool Excel Trick**: place your mouse on the fill handle (the black square at the bottom right corner of the first formula), and when you see the thin cross-hair, double click the mouse. Excel will "auto-copy" the formula all the way down until it finds a blank space in the column to the right of the formula column. This will save you from having to manually copy the formula!
- **Click the Office Button**, select **Excel Options**, click on the **Formulas Tab**.
- Click to **enter a checkmark in the "Manual"** choice for the Formulas Tab dialog box (this will de-select Automatic). This HAS to be done **before** sorting!
- Click **OK**. You can now scroll through your list to view all dups before sorting.
- Click on any cell in the column with the formula in it (Dups) **and Sort** by that field.
- All the duplicates will be together – Delete if desired.

The screenshot shows an Excel spreadsheet with columns A through F. Column A is labeled 'DUPS' and contains formulas: =IF(B2=B1, "DUP", ""), =IF(B3=B2, "DUP", ""), =IF(B4=B3, "DUP", ""), =IF(B5=B4, "DUP", ""), =IF(B6=B5, "DUP", ""), =IF(B7=B6, "DUP", ""), =IF(B8=B7, "DUP", ""), =IF(B9=B8, "DUP", ""), =IF(B10=B9, "DUP", ""), =IF(B11=B10, "DUP", ""), =IF(B12=B11, "DUP", ""), =IF(B13=B12, "DUP", ""), =IF(B14=B13, "DUP", ""), =IF(B15=B14, "DUP", ""), =IF(B16=B15, "DUP", ""), =IF(B17=B16, "DUP", ""), =IF(B18=B17, "DUP", ""), =IF(B19=B18, "DUP", ""), =IF(B20=B19, "DUP", ""), =IF(B21=B20, "DUP", ""), =IF(B22=B21, "DUP", ""), =IF(B23=B22, "DUP", "").

The Excel Options dialog box is open to the Formulas tab. Under 'Calculation options', 'Manual' is selected. Other options include 'Enable iterative calculation' (unchecked), 'Maximum iterations: 100', 'Maximum change: 0.001', 'Recalculate workbook before saving' (checked), 'Working with formulas' (B1C1 reference style unchecked, Formula AutoComplete checked, Use table names in formulas checked, Use GetPivotData functions for PivotTable references checked), and 'Error Checking' (Enable background error checking checked, Indicate errors using this color: [dropdown], Reset Ignored Errors button).

**Note:** Once you have completed this and deleted your duplicates, it's a good idea to go back to **Office Button, Excel Options** and **change** the calculation back to **Automatic!**