

CATEGORY: 5 Review, Skill set: 5.2 Data Integrity

Skill set: 5.2 Data Integrity

5.2.1 Use IF function to test if cell contents are within defined parameters.

A formula can show different text in a cell depending on the values in other cells. Use this to provide information about the quality of the data in the spreadsheet. Text may be more clearly readable in back-and-white printouts than colours produced by conditional formatting (on page 164).

Suppose that H10 contains a total across and J10 contains a total down a table of numbers. You can add a check formula like this:

```
=IF(H10=J10, "OK", "Out of balance!")
```

This 'sanity-check' message shows when C8 is more than 1000:

```
=IF(C8>1000, "Too Large?", "")
```

This message appears when the current date is later than a date entered into D9:

```
=IF(NOW()>D9, "After review date, check assumptions", "")
```

This message in C9 prompts the user that the cell to the right needs a value:

```
=IF(ISNUMBER(D9), "", "Enter the ") & "Review date:"
```

For input cells, it's better to put in Data Validation (on page 166) which is more obvious to the user. For calculation result cells, you can still apply Data Validation, but it will not pop up any messages as it does during data entry. The auditing toolbar provides a facility to draw red circles around invalid calculated results.

Think of other checks that are applicable in your domain. For example, for bar codes and credit card numbers, test the check digits to catch keystroke transposition errors. There are standard methods provided by barcode suppliers.

Data Integrity and Completeness

A useful technique is to add a *health-check* section to a spreadsheet and include it in every printout to provide a degree of assurance to the readers. This requires a cascade of checks as follows:

- 1) Individual cells have a checking cell beside them that tests whether there is an input where there should be, whether the data is of the right type (eg date, text or number), and whether it is in a reasonable scale of values. These cells should return 0 if the data is correct and 1 if there is something that needs checking.
- 2) Rows, columns, and worksheets should show a total of these checking cells. If any of them are non zero, this total cell will also show a non zero result. This total should be in a prominent position in the print area and can be in a distinctive font size, border, shading, or colour. It helps if it is in the same position on each sheet (eg cell C1) so you can quickly check all sheets by repeated Ctrl+PgDn and Ctrl+PgUp.
- 3) The summary sheet of every report, for example the front page of the workbook, should have a grand total check of all the sheet total checks. If this is not zero, you know that somewhere in this workbook a check has failed, so you can then check each sheet in turn to find it.

Compare Worksheets or ranges

Excel 2003 has a Compare Worksheets feature, which you use by opening two workbooks and then choosing the Compare Side by Side option from the Window menu. Excel stacks the sheets vertically and synchronizes them so moving around in one worksheet scrolls the other, letting you compare their contents visually.

Show numeric differences between two worksheets

This procedure helps you to quickly find differences when there are many thousands of cells and it is too easy to overlook differences by a simple visual inspection.

1. Insert a new worksheet to hold the comparison.
2. Select the first range. To select an entire worksheet, select cell A1, then press Shift+Ctrl+End to select all cells to the bottom right-hand corner.
3. Copy the first range (Ctrl+C), switch to the new sheet, select Edit > Paste Special, under Paste select Values, click OK.

4. Select and copy the second range, switch to the new sheet, select **E**dit > **P**aste **S**pecial, select **V**alues, under **O**peration select **S**ubtract, click OK. Press Escape to cancel the *Select destination* message on the status bar. This sheet now contains all the differences between the two sheets.
5. To clear out all the cells containing just a zero, first select the whole worksheet – it should be still selected from step 4. Use **E**dit > **R**eplace; in the Find what box enter a zero; click the **O**ptions button and select **M**atch entire cell contents; click **R**eplace **A**ll and click **C**lose.
6. To select all the cells where any difference has been found, use **E**dit > **G**o to, click **S**pecial, select **C**onstants, and click OK. Use the Tab key to move around the selected cells.

Show any differences between two worksheets

1. Note the bottom right-hand cell in one sheet - the larger range if there is one. Insert a new worksheet to hold the comparison. Go to that cell (eg X123) on the blank sheet. Enter this formula (assume the sheets are Sheet1 and Sheet2):

```
=IF(Sheet1!X123=Sheet2!X123,0,CHAR(39)&Sheet1!X123&CHAR(10)&Sheet2!X123)
```

2. In that formula, the CHAR(39) is an apostrophe which is the label prefix; and CHAR(10) is a newline character. If any pairs of cells are different, this formula will show the Sheet1 value and directly under it the Sheet2 value.
3. Copy that formula into the entire worksheet. This is most simply done by Ctrl+C, Ctrl+Shift+Home, Ctrl+V.
4. Convert all these formulas to value: Ctrl+C, **E**dit > **P**aste **S**pecial > **V**alues.
5. Proceed with steps 5-7 above to remove the zeros and find the differences.

5.2.1 Check your knowledge (answers on page 209)

If you have not already done so, download the sample files from the support website (address on page 4). Open the file **S5test.xls**. On the sheet *Budget08*, enter a formula in cell A1 that displays 'Budget for 2008' if the figure in N70 equals that in N69, and displays 'Budget does not balance' otherwise.