

**General Principles:
Your Data May Change!
Ask The Right Question!**

Copyright 2009
Peter R. Douglass
All rights reserved

Assume Data May Change!

- It is best practice when creating spreadsheets or writing database queries to assume that the given data for a problem might change.

Don't Duplicate Given Data!

- Since the given data for a problem may change, this data should appear as given data in one place only.
- If data given for a problem is repeated in many places, then changing that data will require changes many places.
 - This can be a source of errors as well as a source of unnecessary work.

Ask the right question!

- It is best practice to write your spreadsheets or database queries in such a way that they give correct results for given data, regardless of what that data might be, and not merely the correct results for a particular data set.
- If the data changes, the answer should be correct for the new data set.

A Trivial Example

- Here is a sample problem to illustrate the concept of “Ask the Right Question”
- You are asked to find the sum of the following numbers: 2, 5, 7, 8

Bad Solution

The screenshot shows the Microsoft Excel interface. The title bar reads "Book1 - Microsoft Excel". The ribbon is set to "Home", with the "Editing" group selected. The formula bar shows "A6" and "22". The spreadsheet grid shows the following data:

	A	B	C	D	E	F
1	2					
2	5					
3	7					
4	8					
5						
6	22					
7						
8						

The status bar at the bottom shows "Ready", "Sheet1", "Sheet2", "Sheet3", and "100%".

What's wrong?

- Cell A6 has the correct answer, i.e. 22
- However, notice the formula box contains **the number 22**.
- This means that the answer 22 is “hard coded” in the selected cell (A6).
- If one of the given data cells changed, say A1 was changed from 2 to 3, then the spreadsheet would look like this:

Changed Data / Incorrect Answer

The screenshot shows the Microsoft Excel interface. The title bar reads "Book1 - Microsoft Excel". The ribbon is set to "Home", with the "Editing" group selected. The formula bar shows the active cell "A6" containing the value "22". The spreadsheet grid shows the following data in column A:

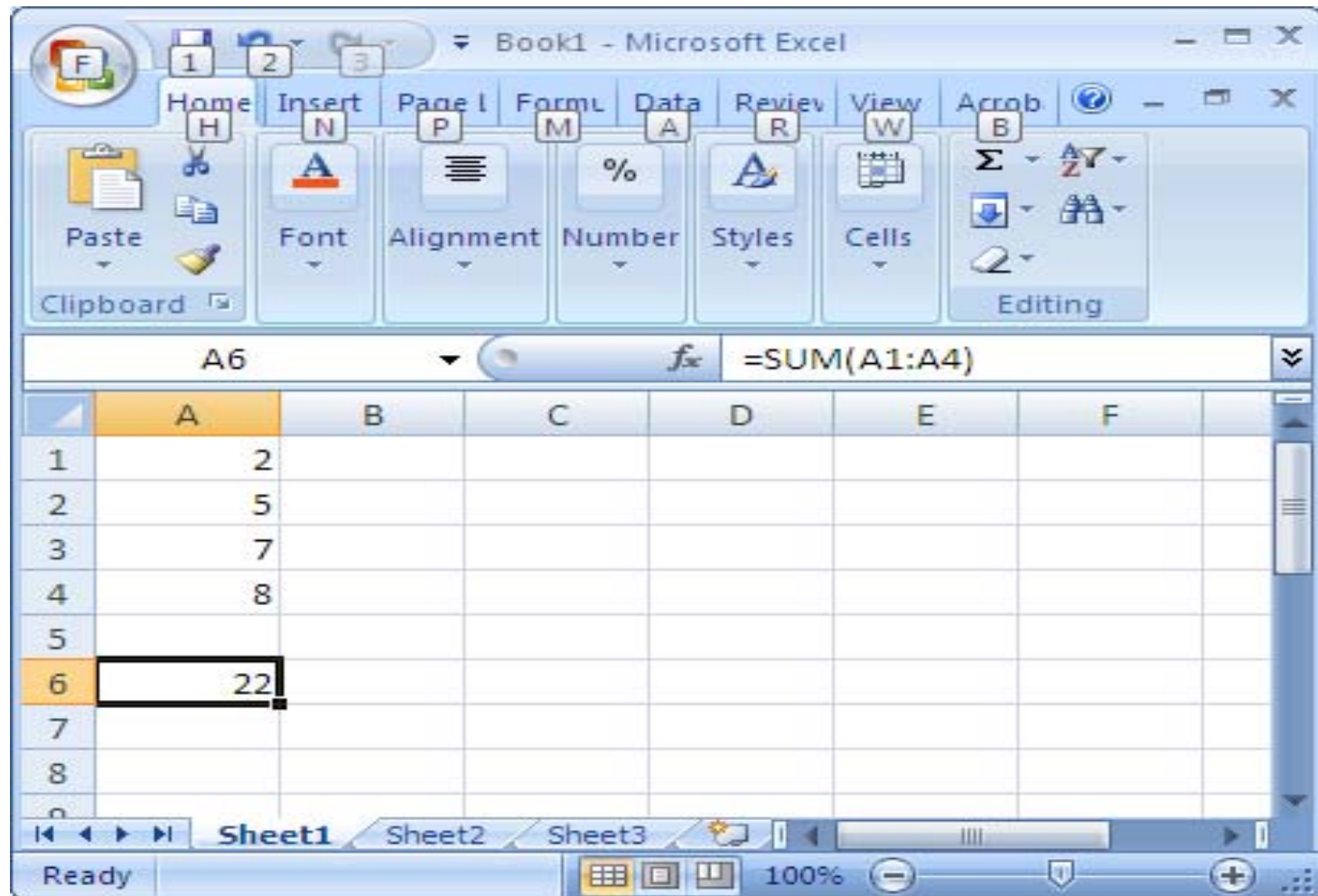
	A	B	C	D	E	F
1	3					
2	5					
3	7					
4	8					
5						
6	22					
7						
8						

The status bar at the bottom indicates "Ready" and "100%".

Changed Data / Incorrect Answer

- The spreadsheet now *appears* to say that the sum of 3, 5, 7 and 8 is 22
- This is the wrong answer, but the error is not obvious without careful inspection.

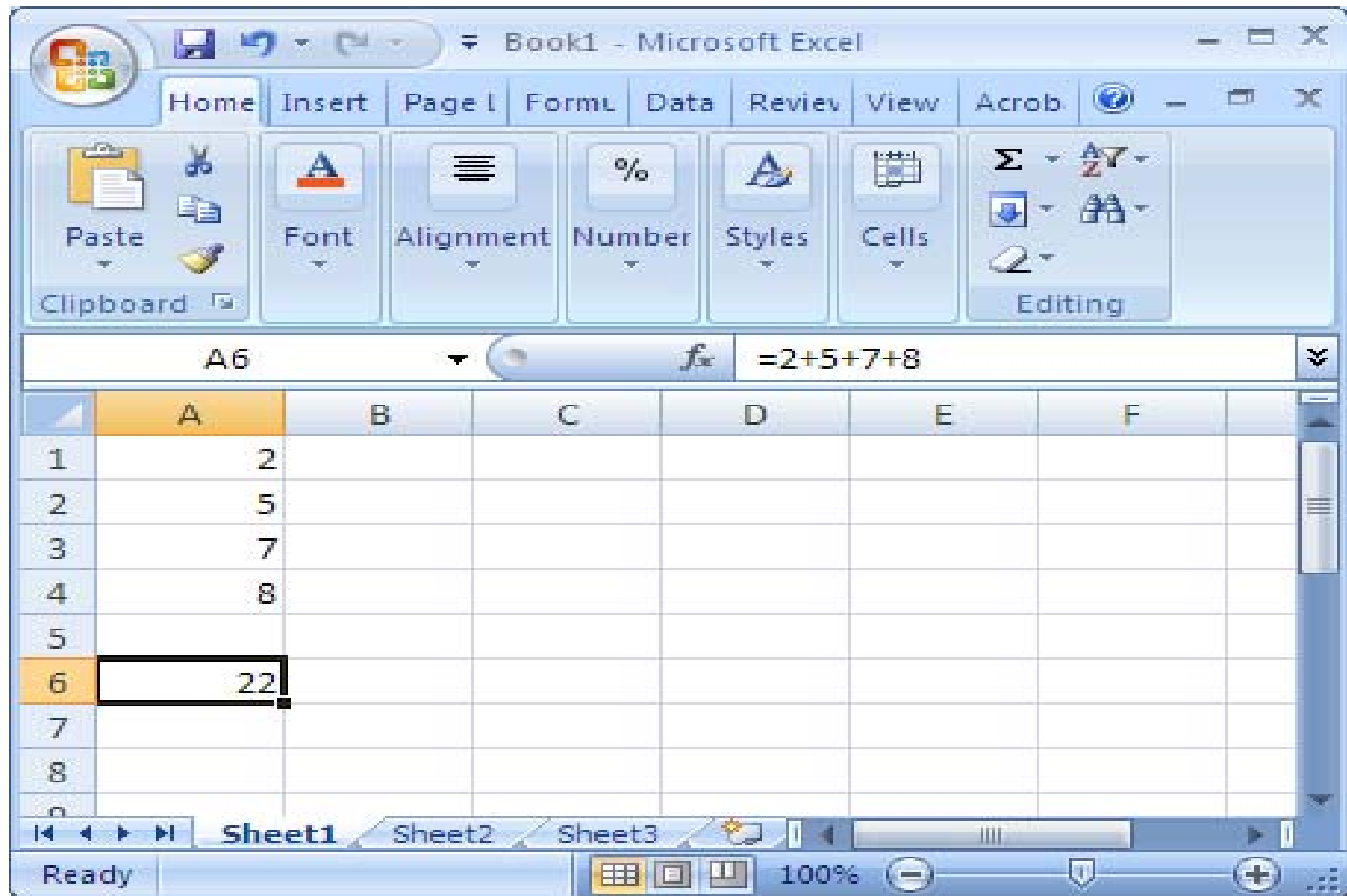
The Problem Solved “Right”



The Fix

- Notice that the formula box now contains a **formula** for calculating the sum.
- If one of the cells containing given data changes, the cell containing the answer will change appropriately.

Another Bad Solution



What's wrong?

- In this example, cell A6 again shows the correct answer.
- Also, the formula box shows that A6 contains a formula for the answer, and not the number 22 itself.
- However, the formula in cell A6 **repeats** the data that appears in cells A1 – A4.
- If value in cell A1 were changed to 3, then the spreadsheet would appear to be saying that the sum of 3, 5, 7 and 8 is 22

A Third Bad Solution

The screenshot shows the Microsoft Excel interface. The title bar reads "Book1 - Microsoft Excel". The ribbon is set to "Home" and includes groups for Clipboard, Font, Alignment, Number, Styles, Cells, and Editing. The active cell is A6, and the formula bar displays the formula $= (A2 + A4 - A1) * A1$. The spreadsheet data is as follows:

	A	B	C	D	E	F
1	2					
2	5					
3	7					
4	8					
5						
6	22					
7						
8						

The status bar at the bottom shows "Ready", "Sheet1", "Sheet2", "Sheet3", and "100%".

What's Wrong?

In this example:

- the cell A6 shows the correct value,
- the cell A6 has a formula not a hard-coded value,
- and that formula does not repeat any data,
- but the formula does not solve the problem of finding the sum of the cells A1 – A4, and only accidentally gives the right answer.