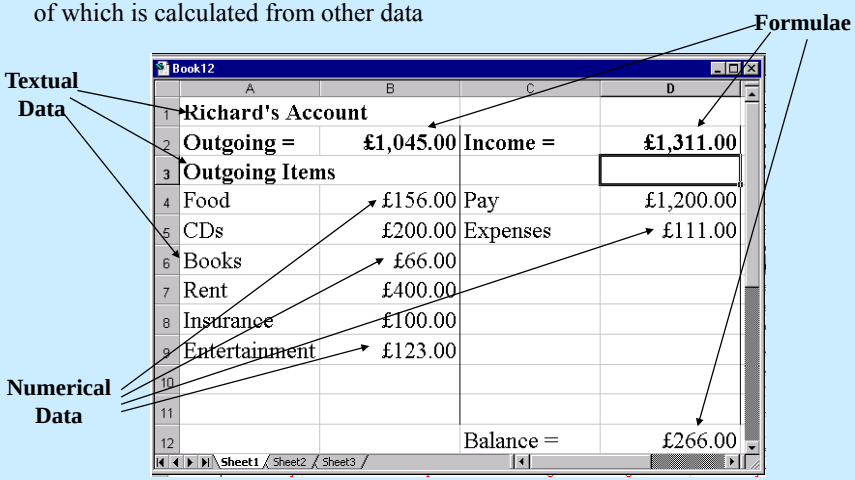
Spreadsheets

A spreadsheet consists of a rectangular arrangement of cells containing data some of which is calculated from other data



Spreadsheet Cells

A cell is identified by its position either: Example in Cell D12

- by **Column and Row Headings** A3, D7, G9 = D2 - B2
  - rows are numbered from 1 upwards
  - columns are labelled A..Z then AA, AB, etc.
- or by **relative position** R[-10]C[-2] means the cell ten rows up and two columns back = R[-10]C - R[-10]C[-2]
- there is a switch to choose which to display

A cell can contain: Examples in previous slide

- **nothing**! - white space is useful for layout purposes to show the contents clearly *B1-D1 – space for the title to flow into*; *D6-D9 – only 2 income items*
- a **number** - one type but multiple formats *B4-B9 and D4-D5*
- a **text string** - useful for labelling adjacent cells *A1, columns A and C*
- a **formula** – which calculates the value (usually from other cells) *B2, D2, D12*

Formatting

The spreadsheet cells can be formatted to highlight various pieces of information

The following can be changed:

- font and font size and style (bold, italic, etc)
- cell size, row height and column width
- text alignment – word wrap, shrink to fit, flow into next cell if blank
- borders
- background pattern, background colour and text colour

Number Representation

There is **only one number type**

but this can be formatted to mean:

- **numbers** as we know them – with different degrees of precision
- **currency** - you can set the unit (£, $, fl, €, δρχ etc)
- **percentages** - use formatting not X/100 since you would have to multiply it again to use it
- **date and time**- integer part is the date (days since 1/1/1900 on PC, 1/1/1904 on Mac) fractional part is the time as the fraction of the day (i.e. 0.5 is 12 noon)
- **floating point** (exponents) - e.g. 7.31E+04 – i.e. $7.31*10^4$ or 73100

Formulae

A formula looks like a text string but starts with "=", e.g.:

= 34 + 12
= A2 + 34
= 23 + B4 + SUM(C1:D13)

The formulae in the cells may refer to other cells:

- in fact, there is a directed acyclic graph of formula dependencies, i.e. each formula means that its cell depends on all the cells in the formula.
- e.g. in the first slide, D12 depends on D2, which depends on D4 and D5 and on B2 which depends on B4, 5, 6, 7, 8 and 9
- There must be no cycles in this collection of dependencies!
  - otherwise the calculation would never stop

Relative and Absolute Addresses I

The identification by row and column name (e.g. =A2) is still a **relative reference**

For instance, if B5 holds "=C7" this means that B5 holds the same value as the cell C7 because cell C7 is:

one to the right (since C is one column after B) and

two down, since 7 is 2 plus 5.

This is important if a formula is moved, either by:

copying and pasting, or - e.g. copying B5 to D8 turns the formula into "=E10" (E is one after D, 10 = 2 plus 8)

inserting a new row - e.g. inserting a new row at the top moves B5 to B6 and the formula becomes "=C8".

Relative and Absolute Addresses II

If you want to fix the reference to a particular cell, then use a "$" in the formula:

= C7 means one to the right and two down

= $C7 means Column C and two down

= C$7 means one to the right and row 7

= $C$7 means Column C, row 7 - and this won't change wherever it is moved to

Use the last of these for constants – e.g. interest rates

since then you can change the interest rate in one place and it will be reflected everywhere

Example

Formula = B5*(1+$F$1)

Constants - $F$1 and $F$2

Base Data Typed In

Book9

	A	B	C	D	E	F
1	Effect of Inflation				Annual Salary Increase =	3.00%
2					Annual Price Increase =	3.50%
3						
4	Year	My Pay	My Costs	Profit		
5	2000	£2,000.00	£1,800.00	£200.00		
6	2001	£2,060.00	£1,863.00	£197.00		
7	2002	£2,121.80	£1,928.21	£193.60		
8	2003	£2,185.45	£1,995.69	£189.76		
9	2004	£2,251.02	£2,065.54	£185.48		
10	2005	£2,318.55	£2,137.84	£180.71		
11	2006	£2,388.10	£2,212.66	£175.45		
12	2007	£2,459.75	£2,290.10	£169.65		
13	2008	£2,533.54	£2,370.26	£163.28		
14	2009	£2,609.55	£2,453.22	£156.33		
15						

Sheet1 Sheet2 Sheet3

Input as .03 and .035 and formatted as percent

Created by copying

Functions

Formulae can include **functions**, whose arguments are either:

- single expressions - 23, A4, 23+A4, etc.
- or ranges of cells - A4:D23 means all of the cells in columns A to D and rows 4 to 23.

There are many functions available in Excel:

- mathematical functions - abs, sin, sqrt
- aggregate functions - sum, avg, count
- text functions - left, right, etc.
- financial, date/time, statistical, etc.

Use of aggregate functions - Cell B2 could have held:

= B4 + B5 + B6 + B7 + B8 + B9 or

= sum(B4:B9)

always use the latter since then we can easily insert new rows and the formula will change automatically.

The IF Function

Some particularly useful functions allow the use of testing and logic

If I don't want to know how much I have overspent, Cell D12 could hold:

=IF(D4>D2, D4-D2, 0) this shows the balance only if it is positive!

The function *IF* takes three parameters:

- a boolean expression D4>D2
- the expression to display if it is true D4-D2
- the expression to display if it is false 0

Compound boolean expressions can be created using the AND, NOT and OR functions, e.g.:

=IF(OR(D4>D2, B5=0), D4-D2, 0) display balance if positive or I haven't spent anything on CD's

Some Special Functions – COUNTIF and SUMIF

COUNTIF takes a range of cells and counts the number which match a criterion:

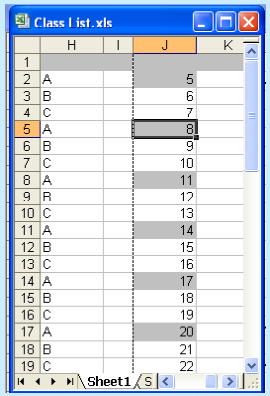
=COUNTIF($H$2:$H$89,"A") counts the cells in column H from rows 2 to 89 which hold the string "A"

SUMIF takes one range, a criterion value and a second range & returns the sum of the cells in the second range, for which the first range match the criterion:

=SUMIF($H$2:$H$89, "A", $J$2:$J$89) adds the cells in column J from rows 2 to 89 which are in the same row as cells of column H holding "A"

This only sums the grey cells

There is also *COUNTBLANK*



LOOKUP Functions I

Note use dollars to indicate the use of the same (constant) area

The lookup functions can be used, for instance, to turn a score into a band as shown on the left

- Column N holds a set of scores out of 50
- Q4 to R24 hold a translation table – e.g. 35-39 is A3

Column O contains the calculated band using the *VLOOKUP* function and for O6 is

=VLOOKUP(N6, $Q$4:$R$24, 2)

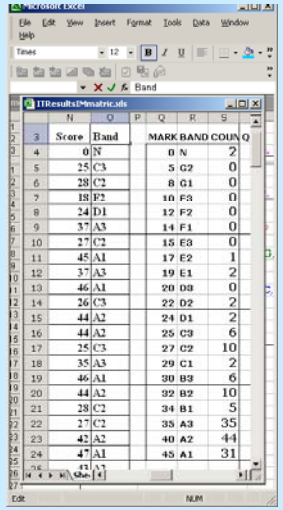
in which

N6 holds the score which is to the right of the band

Q:$R24 holds the table – note the dollars to ensure that it can be copied down

2 means the band is in column 2 of the table

This is called a **range lookup**



LOOKUP Functions II

In this example the lookup functions can be used to produce an **exact lookup**

- which is good if you're matching non-numerical data

We add a fourth boolean parameter which

if TRUE (the default) asserts range lookup

if FALSE asserts exact lookup, e.g.

=VLOOKUP(O6, $E$3:$F$25, 2, FALSE)

in which

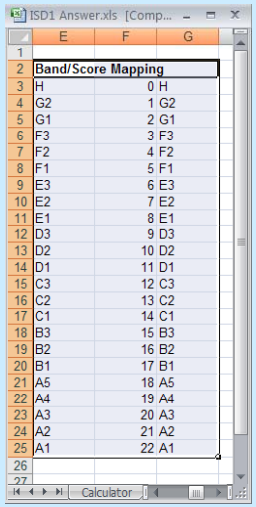
O6 from the last slide holds the band, "C2"

E:$R25 holds the table – note the dollars again

2 means the band score is in column 2 of the table

it's exact lookup so the "C2" will give the result 13

but "C4" would give an error



There is also *HLOOKUP*

Paste Special

In Excel, the Paste section of the Home tab is very important

If you do Copy and Paste then the cell is copied in total

- i.e. the formatting is copied, as well as the formula (with the references modified as shown on Slides 38-39)

The Paste options and, in particular, Paste Special allows you to do any of:

- Copy just the format
- Copy the current value of a formula rather than the formula itself
- Transpose a column into a row and vice versa
- Use the number in the copied cell to operate on the target cell – e.g. add to it, multiple by it etc.
- This is the way you turn an active hyperlink into a piece of text
  - You put the number 1 into a cell, copy it and then multiply the hyperlink by it

Further Aspects of Spreadsheets I

Sorting and Searching

- Select a rectangular area to sort
- Sort by rows or columns
- If by rows, choose column(s) to hold the sort key
- Search-and-replace as for word processors

Charts

- A chart is a graphical display of the data in a spreadsheet
- *Insert:Chart* allows you to visualise the data as a graph, a scatterplot, a pie chart and so on
- Typically you select a rectangular area of the grid and then generate the chart based on this data
- The first column or row may be used specially, e.g. as a label or the X-axis of a scatterplot

Further Aspects of Spreadsheets II

Auditing

- *Formula Auditing* in the *Formulas* tab allows you to show the dependencies between cells by displaying linking lines

Macros

- A macro is a stored set of spreadsheet operations
- You can record a series of operations in a similar way to a tape recorder
- The macro can the be rerun whenever you like

Other OLE Functions

- In a similar manner to Word, Excel allows you put other kinds of document in a cell – e.g. WP documents, drawings and so on

Further Aspects of Spreadsheets III

Database Functions

- The *Data* tab holds a number of functions for manipulating a series of rows as a set of records
- The top row of the series can be taken to be column names - e.g. for charts, mail merge, etc.
- For instance, you can sort the set, filter it based on a conditions, and so on

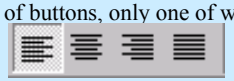
Visual Basic

- Finally, a spreadsheet can be extended with programming in Visual Basic
- Many of the features above generate VB automatically

User Interface Features of Excel

The Excel user interface is organised using features which have become standard in Mac/Windows programs. Here are some of them:

- A **menu bar** – this holds a quick method for accessing the features
- **Tool bars** – iconic representations of functions – there are very many of these in Excel and the *View:Toolbars* dialogue controls which are visible
- **Checkboxes** – these will control a function by checking or unchecking the box e.g. emboldening happens only if the "B" box is checked
- **Radio buttons** – a set of buttons, only one of which can be active at any time



e.g. – these control justification of the cells and only one can be active

- **Combo boxes** – these provide a text box and a drop down menu

- **Tool tips** – are displayed as you move over the various features of the spreadsheets

Using Java or XHTML you can program these in any application you write with relative ease

The HCI Style of Excel

There are several styles of interaction supported by software

- **command language** – the user specifies the commands by typing in sentences in a particular language – e.g. SQL for databases
- **form filling** – data entry often precedes by having the user fill in a form
- **indirect manipulation** – the user manipulates some part of the data which affects another part
- **direct manipulation** – the user directly edits the data representation

Excel provides indirect and direct manipulation:

editing the Formula Bar ultimately places a value in a cell and hence is **indirect manipulation**

editing directly in a cell is **direct manipulation** as is selecting cells by point-and-click when building up formulae

Pivot Tables I

A **Pivot Table** is a table that summarises data from a set of data records by analysing the distribution of one field with respect to others.

Example:

The Data

Account		
Item	**Method**	**Amount**
CD	CrdtCd	£12
Book	CrdtCd	£10
Food	DbtCd	£15
Cats	Cash	£5
Food	DbtCd	£3
Car	CrdtCd	£25

Sum of Amount	**Method**			
Item	**Cash**	**CrdtCd**	**DbtCd**	**Total**
Book		£10		£10
Car		£25		£25
Cats	£5			£5
CD		£12		£12
Food			£18	£18
Total	£5	£47	£18	£70

The Pivot Table

Pivot Tables II

In the example six data records have been entered in the first three columns

Each record represents one purchase in terms of the kind of item bought, the payment method and the amount paid

The pivot table on the right sums the payments for each combination of item type and payment method

A pivot table is a technique drawn from **On-Line Analytical Processing**, which is an emerging methodology for taking a set of data and looking at it in various ways

Use of Pivot Tables

To use the technique in Excel, you must:

1. Select a set of data including the headings
2. Choose the menu item *Data:Pivot Table and Pivot Chart Report*
3. Choose a structure – in the example, there are column and row values, summed values in the cells and totals are also added
4. Choose where to put the pivot table
5. Choose the formula used – in this case summing the values
6. Choose what goes where. In this case, *Item* is the row, *Method* the column and *Amount* the summed values in the cells

You can also create **Pivot Charts**

Uses of Spreadsheets I

Timetables

- The rectangular grid gives a structure appropriate to display a timetable
- The data and time formatting supports the correct kinds of annotation

Customised Document Preparation

- The **mail merge** Word Processor function supports the automatic creation of multiple customised copies of documents such as forms or mail shots
- The WP document holds placeholders for data from the spreadsheet – e.g. in Word the occurrence of "<<Name>>" will be replaced by a name from the spreadsheet. To do this:
  - place a series of headings in row 1, including "Name";
  - place each record in a separate row, with the person's name in the column headed "Name"
  - the Mail Merge command will produce one printout for each row in the Spreadsheet

Uses of Spreadsheets II

Accounts

- Book-keeping layout also fits the grid structure
- Totalling functions of reasonable complexity are highly appropriate

Planning

- The immediacy of calculation allows the effect of various values for parameters to be checked quickly
- Thus "what-if" planning can be carried out
- Plans and actual occurrences can be contrasted
- The differences can be charted

Uses of Spreadsheets III

Data Analysis

- A tabular layout of results is well supported – see pivot table slide.
- The availability of mathematical and statistical functions makes this easier …
  … but calculation errors must be avoided.
- Charts and graphs can be drawn easily.

Simple Data Management

- The rectangular grid can be used to hold simple records – one row per record.
- The Edit:Find command can be used for simple data retrieval. Data:Sort puts the data in order.
- Summaries can be calculated easily.
  … but for complex data structures you probably need to go to a database system.

Errors in Spreadsheets I

Using a computer for data analysis and presentation is as error-prone as any human activity

Here are some of the sources of error

data collection errors – ensure data is reliably collected and check if not sure

- **duplicating** data collection may help
- **automatic** data entry – e.g. using a bar code reader – improves matters

data entry errors – keying in data is tedious and error prone

- use automatic checks
- avoid repetitious data entry –
  e.g. use **combo boxes** to avoid re-entry – e.g. for job titles
  auto-completion may help for repeated entries in the same column - this turns the first few characters into a previous entry
  provide a supportive interface for data entry
- make automatic links to other data sources rather than re-enter – at worst use copy-and-paste

Errors in Spreadsheets II

algorithmic errors – it is easy to get the formulae wrong!

- check the results for representative data sets
  - values from the centre and extremes of a range
- summarise and plot the data to check for obvious errors – e.g. a value wildly out of range

numeric errors – remember that the computer has limited memory – see next slide

software errors – professional software contains bugs too

hardware errors – these occur very rarely, but they do happen

misinterpretation of results – it is easy to misread data

- poor use of formatting can result in all sorts of misinterpretation error
- e.g. parallax errors – use regularly spaced lines along long rows

Numeric Errors – Trying to Represent the Infinite - I

One main problem with computer representation of numbers is the **finiteness** of the computer's representational ability.

Numerical algorithms such as those used in spreadsheets arise as a computerised form of mathematical functions and processes, but there is a big difference:

- most mathematical systems deal with **infinite sets of numbers**
- e.g. normal **integer arithmetic** deals well with all of the integers –
  the system works just as well for 1 + 1234567890123456789 as it does for 1+1
  but the set of integers is (countably) **infinite**
- similarly **real number** arithmetic can handle any combination of numbers of any size and precision, even though there are even more of these than there are integers -
  real numbers form an (uncountably) **infinite** set

Numeric Errors – Trying to Represent the Infinite - II

Computers can only deal with **a finite subset** of these infinite sets.

- choosing which sets to represent numbers with is dealt with in Computer Systems

Here the important issue is that the result of any calculation might lie outside the range of the computer's represented set of numbers

Recall from Systems and Networks that with n bits we can represent 2^n different integers

- we usually choose the values $-(2^{n-1}-1)$ to $+(2^{n-1}-1)$
- if n=16 then this means –32,767 to +32,767

One problem comes when a calculation goes beyond that range:

- e.g. adding 30,000 and 20,000
- this should give 50,000 but that is out of range

Numeric Errors – Trying to Represent the Infinite - III

A second problem comes when you combine real numbers of very different sizes:

- we use the floating point technique to represent real numbers –
  i.e. we turn a real number, N say, into two parts –
  a number, M, between 0.1 and 1 and
  the number of times, D, we must multiple or divide by two to get our number
  i.e. $N = 2^D * M$
- this allows you to represent both 12345670000 and .00007654321
- but if we try adding them, then the small number gets lost
- the effect of this can be severe if we are combining a lot of numbers

The Kinds of Error that Arise

The main problems are:

- **overflow** - a calculation produces a value out of range
  e.g. adding 30,000 and 20,000
- **rounding** - important parts of the result of a calculation are lost since only a restricted precision of each subsidiary calculation is kept..
  e.g. adding 12345670000 and .00007654321
- **cancellation errors** – the subtraction of two very nearly equal numbers resulting in a value of little precision
  e.g. calculating sin(22)

The Importance of Numerical Analysis

If you are using a spreadsheet for complex data manipulation, it is important to use algorithms which are guaranteed to avoid rounding and overflow errors

Numerical analysis is the study of such algorithms

One example of the importance of this:

- In the 1970s, a series of box girder bridges collapsed across British motorways
- Numerical algorithm errors were the cause of this
- One expert found that more than 50% of the engineering calculations he examined were incorrect to the second significant figure