

Calculations in Excel

Spreadsheet formulas
Functions

Spreadsheet formulas

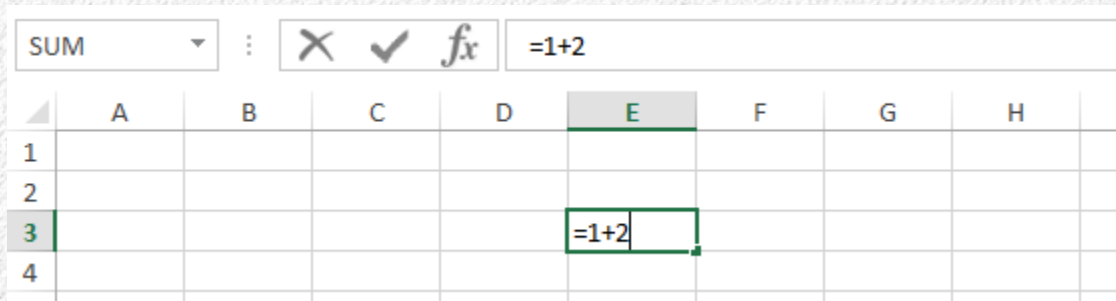
- Spreadsheets can be used to do a wide range of calculations, data analysis
 - Arithmetic
 - Built-in functions
- Functions can be “nested” and combined
- This is the reason to use spreadsheets

Cell formulas: basic arithmetic

- Use an “=” to start a formula
- Arithmetic operators are:

| Operation | Symbol |
|----------------|--------|
| Addition | + |
| Subtraction | - |
| Multiplication | * |
| Division | / |
| Exponent | ^ |

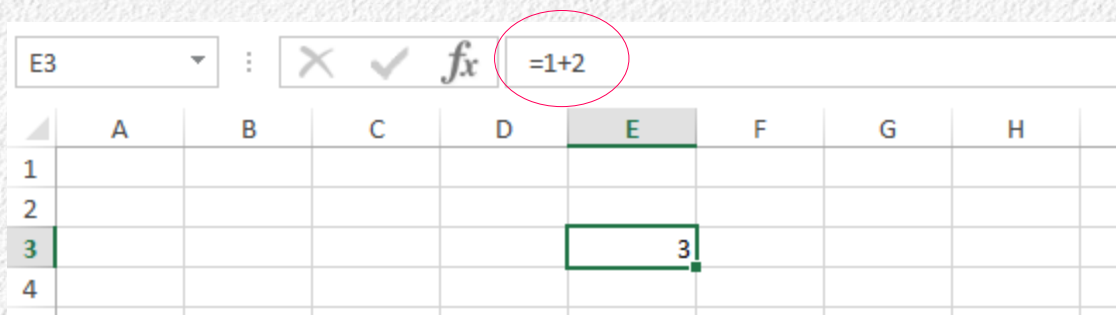
Entering formulas, getting the results



A screenshot of the Microsoft Excel interface. The formula bar at the top shows the text '=1+2'. Below it, a grid of cells is visible. Cell E3 is selected and contains the text '=1+2'. The formula bar also displays '=1+2'. The grid shows columns A through H and rows 1 through 4.

| | A | B | C | D | E | F | G | H |
|---|---|---|---|---|------|---|---|---|
| 1 | | | | | | | | |
| 2 | | | | | | | | |
| 3 | | | | | =1+2 | | | |
| 4 | | | | | | | | |

Editing a cell – the formula bar and the cell itself show the formula



A screenshot of the Microsoft Excel interface. The formula bar at the top shows the text '=1+2', which is circled in red. Below it, a grid of cells is visible. Cell E3 is selected and contains the number '3'. The formula bar still displays '=1+2'. The grid shows columns A through H and rows 1 through 4.

| | A | B | C | D | E | F | G | H |
|---|---|---|---|---|---|---|---|---|
| 1 | | | | | | | | |
| 2 | | | | | | | | |
| 3 | | | | | 3 | | | |
| 4 | | | | | | | | |

After hitting Enter, the formula bar still shows the formula, but the cell shows the result

Exponentials

$=3^2$ gives 9

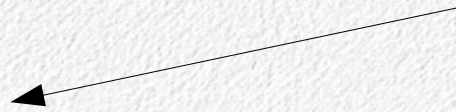
What is $9^{(1/2)}$?

What is the cube root of 8?

| B | C |
|--------------|----------|
| Formula | Result |
| $=3^2$ | 9 |
| $=9^{1/2}$ | 4.5 |
| $=9^{(1/2)}$ | 3 |
| $=9^{0.5}$ | 3 |
| $=9^{-0.5}$ | 0.333333 |
| $=8^{(1/3)}$ | 2 |

Parentheses used to determine order of operations

Operations inside parentheses done first



Comparison operators

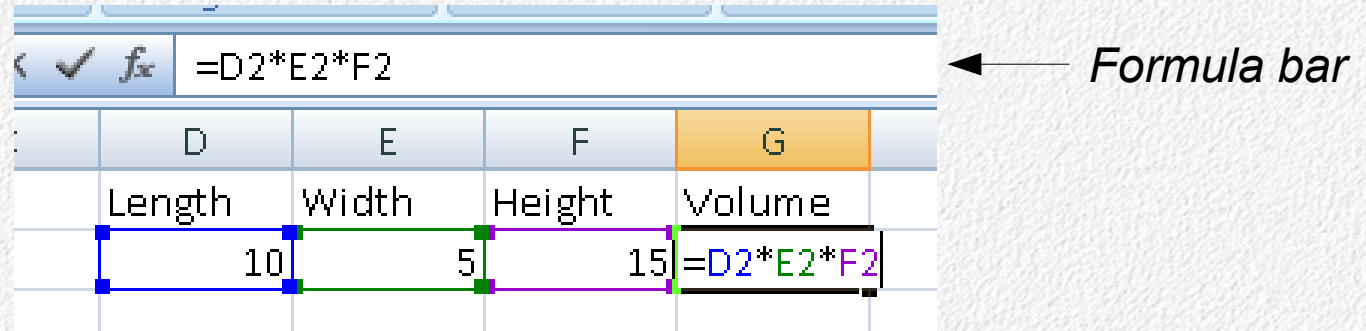
- Used to compare one cell to another
- Return only “True” or “False” (true is equal to 1 to a computer, false is equal to 0)
- = , equals
- > , greater than
- < , less than
- >= , greater than or equal to
- <= , less than or equal to
- <> , not equal to

| F | G |
|---------|--------|
| Formula | Result |
| =1=0 | FALSE |
| =1>0 | TRUE |
| =1<0 | FALSE |
| =1>=0 | TRUE |
| =1<=0 | FALSE |
| =1<>0 | TRUE |

Cell references

- Spreadsheet formula in one cell often uses data in another cell(s)
 - If the data changes, the formula automatically re-calculates, updates formula's result
- Pointers to other cells are cell references
- Cell references can be absolute or relative
 - Absolute = refers to a particular cell, won't change if the formula is copy/pasted elsewhere
 - Relative = refers to a cell by its position relative to the cell that holds the formula, changes if the formula is copy/pasted elsewhere

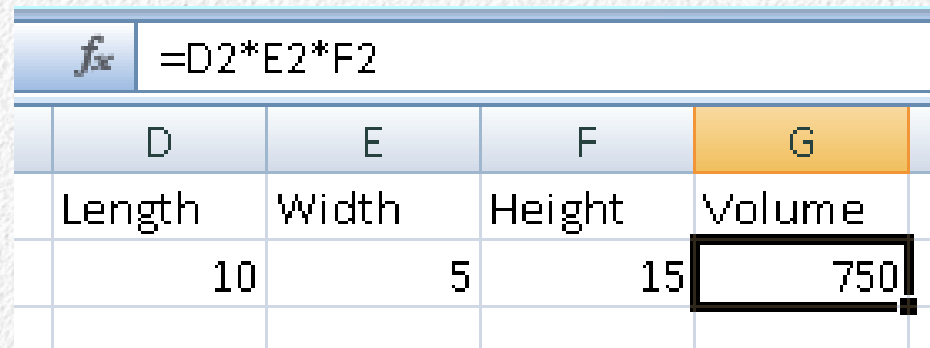
Example: relative references



The screenshot shows the Excel interface. The formula bar at the top contains the formula `=D2*E2*F2`. Below it is a table with columns D, E, F, and G. Row 2 contains the labels Length, Width, Height, and Volume. Row 3 contains the values 10, 5, 15, and the formula `=D2*E2*F2`. Colored boxes highlight the relative references: a blue box around D2 (10), a green box around E2 (5), and a purple box around F2 (15). An arrow points from the text "Formula bar" to the formula bar.

| | D | E | F | G |
|--|--------|-------|--------|------------------------|
| | Length | Width | Height | Volume |
| | 10 | 5 | 15 | <code>=D2*E2*F2</code> |

`=D2*E2*F2` says “the cell in my row three columns left, the cell in my row two columns left, the cell in my row one column left”



The screenshot shows the same Excel interface as above, but the formula bar now only shows the formula `=D2*E2*F2` and the cell G2 now displays the result of the calculation, 750.

| | D | E | F | G |
|--|--------|-------|--------|--------|
| | Length | Width | Height | Volume |
| | 10 | 5 | 15 | 750 |

Once entered the formula bar still shows the formula, but the cell shows the result of the calculation

Relative references change when the cell is copied and pasted

fx =D2*E2*F2

| D | E | F | G |
|--------|-------|--------|-----------|
| Length | Width | Height | Volume |
| 10 | 5 | 15 | =D2*E2*F2 |
| 8 | 6 | 5 | |

fx =D3*E3*F3

| D | E | F | G |
|--------|-------|--------|--------|
| Length | Width | Height | Volume |
| 10 | 5 | 15 | 750 |
| 8 | 6 | 5 | 240 |

Absolute references

- References that don't depend on position of the formula
 - Don't change when cell is copied/pasted
 - Made by placing a dollar sign before the row and/or column reference
 - Can mix them – e.g. absolute column, relative row
- Example: calculating relative frequencies
 - Have a count for each row, and a total
 - Need to divide the count for each row by the same total
- Need absolute cell references allow the count to change with each row, but keep the total the same

Allele frequencies – proportions

First entry

| fx =E2/E\$8 | | | | |
|-------------|--------|-------|-----------|---|
| C | D | E | F | G |
| | Allele | Count | Frequency | |
| | A | 22 | 0.12941 | |
| | B | 135 | | |
| | C | 9 | | |
| | D | 3 | | |
| | E | 1 | | |
| | | | | |
| | Total | 170 | | |
| | | | | |

Copied and pasted to rest

| fx =E6/E\$8 | | | | |
|-------------|--------|-------|-----------|---|
| C | D | E | F | G |
| | Allele | Count | Frequency | |
| | A | 22 | 0.12941 | |
| | B | 135 | 0.79412 | |
| | C | 9 | 0.05294 | |
| | D | 3 | 0.01765 | |
| | E | 1 | 0.00588 | |
| | | | | |
| | Total | 170 | | |
| | | | | |

Why copy/paste instead of entering the formula repeatedly?

Functions

- Functions are mini programs built into Excel
 - Most take one or more arguments
 - All return a result
- Structure of functions:
 - Function name
 - Followed by open parentheses with no space
 - Arguments (if needed) are within the parentheses
 - Spaces within the parentheses are ignored

Examples of functions with no arguments

| | A | B | C |
|---|----------|-----------------|--|
| 1 | Function | Returns | Explanation |
| 2 | pi() | 3.141592654 | Value of pi to 15 decimal places |
| 3 | rand() | 0.751827802 | A random uniform number |
| 4 | false() | FALSE | Enters the logical value FALSE into the cell |
| 5 | true() | TRUE | Enters the logical value TRUE into the cell |
| 6 | today() | 1/17/2012 | Enters today's date |
| 7 | now() | 1/17/2012 10:17 | Enters today's date and current time |
| 8 | | | |

Functions with variable numbers of arguments

- Some functions operate on whatever entries (of the right data type) you specify
 - All arguments treated the same
 - Order/position doesn't matter
- The number of entries varies depending on the data
- Examples are `sum()`, `average()`, `stdev()`...

Sums take variable numbers of arguments

A screenshot of an Excel spreadsheet. The formula bar at the top shows `=SUM(D2,D3,D4,D5)`. The spreadsheet has columns C through I and rows 3 through 6. Cell D2 contains 'Data', D3 contains 9.5, D4 contains 9.9, D5 contains 12.0, and D6 contains 10.7. Each of these cells is highlighted with a different colored border (blue, green, purple, red). In row 7, cell C7 contains 'Sum' and cell D7 contains the formula `=SUM(D2,D3,D4,D5)`. A tooltip below the formula bar shows the syntax: `SUM(number1, [number2], [number3], [number4], [number5], ...)`.

| | C | D | E | F | G | H | I |
|---|-----|--------------------------------|---|---|---|---|---|
| 3 | | Data | | | | | |
| | | 9.5 | | | | | |
| | | 9.9 | | | | | |
| | | 12.0 | | | | | |
| | | 10.7 | | | | | |
| | Sum | <code>=SUM(D2,D3,D4,D5)</code> | | | | | |

A screenshot of an Excel spreadsheet. The formula bar at the top shows `=SUM(D2:D5)`. The spreadsheet has columns C through F and rows 3 through 6. Cell D2 contains 'Data', D3 contains 9.5, D4 contains 9.9, D5 contains 12.0, and D6 contains 10.7. The entire range D2:D5 is highlighted with a blue border. In row 7, cell C7 contains 'Sum' and cell D7 contains the formula `=SUM(D2:D5)`. A tooltip below the formula bar shows the syntax: `SUM(number1, [number2], ...)`.

| | C | D | E | F |
|---|-----|--------------------------|---|---|
| 3 | | Data | | |
| | | 9.5 | | |
| | | 9.9 | | |
| | | 12.0 | | |
| | | 10.7 | | |
| | Sum | <code>=SUM(D2:D5)</code> | | |

Ranges of cells with a colon

Functions with specific arguments

- Some functions take a specific set of arguments
 - Optional arguments = default settings will be used if not specified (may get wrong answer, but will run)
 - Mandatory arguments = you will get an error message without them
- The order/position of the arguments tells Excel what they are – need to enter them in the correct order
- Mandatory first, optional at the end (if any)

Example: the if() function

- The function if() executes an “if...then...else” statement
- It takes three arguments
 - A logical test (if)
 - A value if the test returns true (then)
 - A value if the test returns false (else)
- The order of the arguments tells Excel which is which

| B | C | D | E | F | G | H |
|---|------|--------|---|---|---|---|
| | | Data | Above the mean? | | | |
| | | 9.5 | =if(d2>d\$7) | | | |
| | | 9.9 | IF(logical_test, [value_if_true], [value_if_false]) | | | |
| | | 12 | | | | |
| | | 10.7 | | | | |
| | Mean | 10.525 | | | | |

Excel prompts you for the needed arguments as you enter the function

| B | C | D | E | F | G | H |
|---|------|--------|---|---|---|---|
| | | Data | Above the mean? | | | |
| | | 9.5 | =if(d2>d\$7,"Yes") | | | |
| | | 9.9 | IF(logical_test, [value_if_true], [value_if_false]) | | | |
| | | 12 | | | | |
| | | 10.7 | | | | |
| | Mean | 10.525 | | | | |

| B | C | D | E | F | G | H |
|---|------|--------|---|---|---|---|
| | | Data | Above the mean? | | | |
| | | 9.5 | =if(d2>d\$7,"Yes","No") | | | |
| | | 9.9 | IF(logical_test, [value_if_true], [value_if_false]) | | | |
| | | 12 | | | | |
| | | 10.7 | | | | |
| | Mean | 10.525 | | | | |

The “function wizard”

- There are various ways that Excel guides you in properly using functions
- The “function wizard” is a window with each of the arguments listed, with a brief description of what they mean
- Help can be easily accessed as well

Function wizard example

- The function wizard for if()

The screenshot shows the 'Function Arguments' dialog box for the IF function. The dialog is titled 'Function Arguments' and has a question mark icon and a close button (X) in the top right corner. The function name 'IF' is displayed in the top left. The arguments are listed as follows:

| | | | |
|----------------|--------|---|-------|
| Logical_test | D2>D57 | = | FALSE |
| Value_if_true | "Yes" | = | "Yes" |
| Value_if_false | "No" | = | "No" |

Below the arguments, there is a brief explanation of the function: 'Checks whether a condition is met, and returns one value if TRUE, and another value if FALSE.' A note specifies that 'Logical_test is any value or expression that can be evaluated to TRUE or FALSE.' At the bottom, the 'Formula result' is shown as 'No'. There are 'OK' and 'Cancel' buttons at the bottom right, and a 'Help on this function' link at the bottom left.

Annotations with arrows point to various parts of the dialog:

- Name of the function:** Points to the 'IF' label in the top left.
- Arguments:** Points to the top right corner of the dialog box.
- Values for each argument currently entered in cell:** Points to the 'Yes' and 'No' values in the 'Value_if_true' and 'Value_if_false' fields.
- Brief explanation of function's purpose:** Points to the text 'Checks whether a condition is met, and returns one value if TRUE, and another value if FALSE.'
- Brief explanation of currently selected argument:** Points to the 'Logical_test' field.
- Value returned by the function with the current arguments:** Points to the 'Formula result = No' text.

Combining functions, doing math on their results

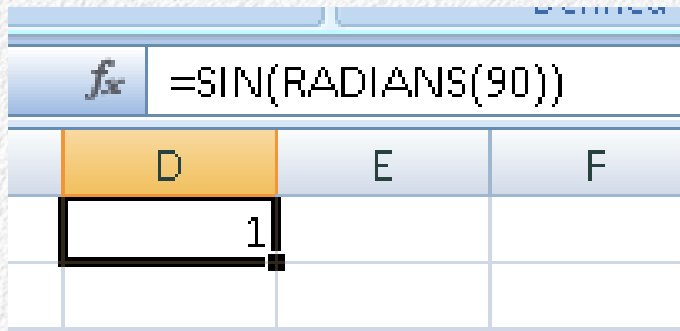
| Alignment | | Number | |
|-----------|--------------------------|--------|--|
| f_x | =SUM(D2:D6)/COUNT(D2:D6) | | |
| D | E | F | |
| Data | | | |
| 11.9 | | | |
| 11.9 | | | |
| 12.8 | | | |
| 15.3 | | | |
| 10.7 | | | |
| 12.52 | | | |

| f_x | | =SUM(D2:D5)^2 | |
|------|-----------------------|---------------|--|
| D | E | | |
| Data | Square of sum of data | | |
| 9.5 | 1772.41 | | |
| 9.9 | | | |
| 12.0 | | | |
| 10.7 | | | |

Nesting functions

- Functions can be used as arguments to other functions
 - Evaluated from inside out
- Examples...

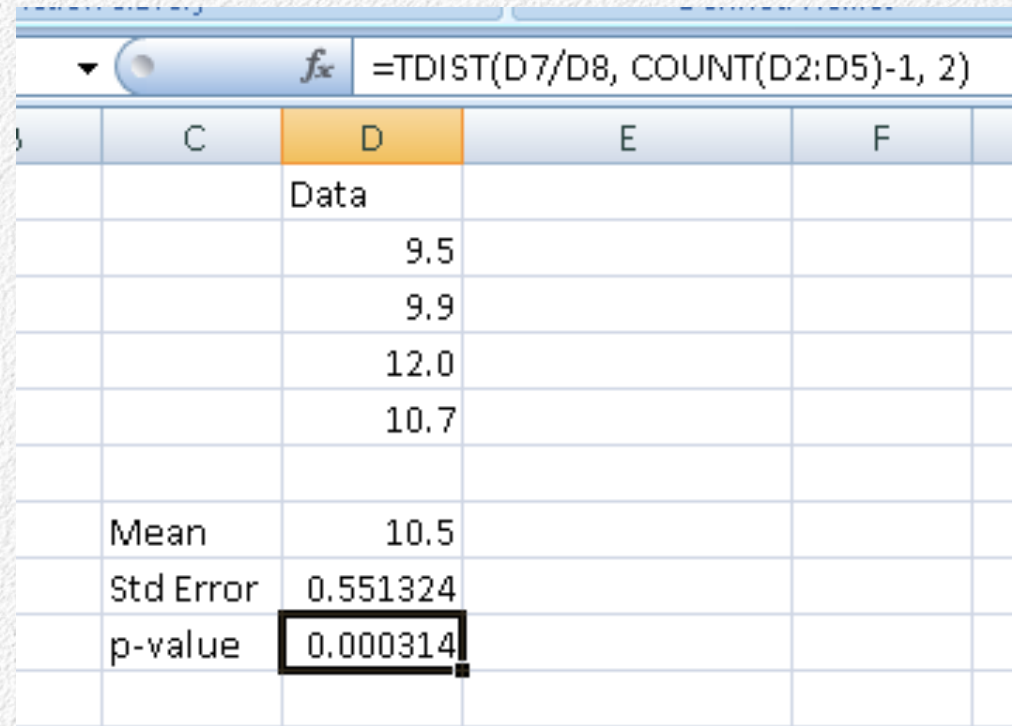
Nesting functions



A screenshot of an Excel spreadsheet. The formula bar at the top shows the formula `=SIN(RADIANS(90))`. Below the formula bar, the spreadsheet grid shows columns D, E, and F. Cell D1 is highlighted with a thick black border and contains the value 1.

| D | E | F |
|---|---|---|
| 1 | | |

But, be careful...
harder to error
check



A screenshot of an Excel spreadsheet. The formula bar at the top shows the formula `=TDIST(D7/D8, COUNT(D2:D5)-1, 2)`. Below the formula bar, the spreadsheet grid shows columns C, D, E, and F. Cell D1 is labeled "Data" and contains a list of values: 9.5, 9.9, 12.0, and 10.7. Cell D6 is labeled "Mean" and contains 10.5. Cell D7 is labeled "Std Error" and contains 0.551324. Cell D8 is labeled "p-value" and contains 0.000314, which is highlighted with a thick black border.

| C | D | E | F |
|---|-----------|----------|---|
| | Data | | |
| | 9.5 | | |
| | 9.9 | | |
| | 12.0 | | |
| | 10.7 | | |
| | | | |
| | Mean | 10.5 | |
| | Std Error | 0.551324 | |
| | p-value | 0.000314 | |

Data types in Excel

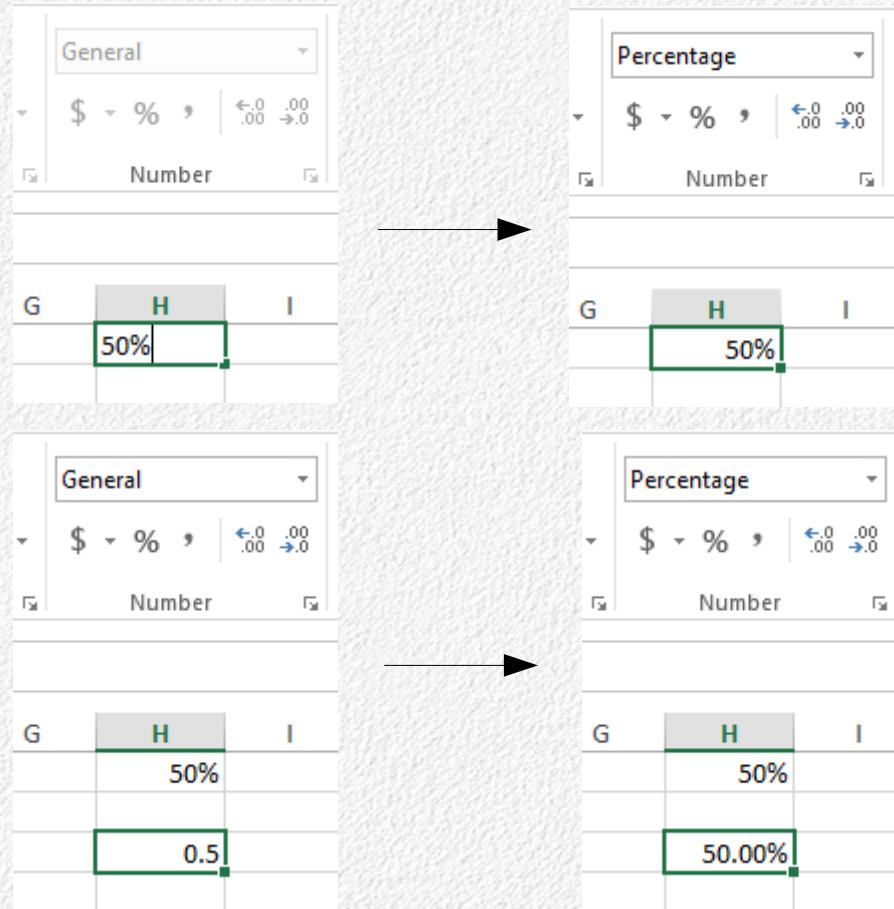
- Unlike Access, we don't have to assign a data type to the data we enter in Excel
- But, Excel does have data types – it assigns them as you enter the data...in other words, it guesses
- Functions, formulas only work on the correct data type
- Common data types to be aware of are:
 - Numbers
 - Percents
 - Text
 - Dates/times
 - Logical

Numbers

- All numbers are stored as double-precision floating point values
 - Why? Integer math drops decimal places
 - If you enter an integer, it will display in the cell without decimal places, but internally it's recorded as floating point
- Numbers displayed right-justified by default
- Most of the time number entered and number stored and displayed are the same, aside from rounding
- Exception: percents

Percents are stored as proportions

- Entering 50% causes Excel to:
 - Store the value 0.5
 - Set the cell type to Percentage
 - Display 50%
- Entering 0.5 and then setting the display type to Percentage gives same result



Setting the display type to percent before entering data

- Setting the cell type to Percentage and then entering 0.5 gives 0.5%
- Changing this back to General shows the number stored is 0.005

Percentage

\$ % , ←.0 .00 →.0

Number

| G | H | I |
|---|--------|---|
| | 50% | |
| | 50.00% | |
| | | |

Percentage

\$ % , ←.0 .00 →.0

Number

| G | H | I |
|---|--------|---|
| | 0.50% | |
| | 50.00% | |
| | | |

General

\$ % , ←.0 .00 →.0

Number

| G | H | I |
|---|--------|---|
| | 50% | |
| | 50.00% | |
| | 0.005 | |

Text

- Any non-numeric entries are stored as text
 - Mixes of numbers and letters (24 g) stored as text
- Left-justified by default
- Text entries won't be used in calculations, but don't cause error messages
 - Text skipped over
- Various text functions
 - `len()` - length of a text string
 - `lower()` - converts to lower case
 - `proper()` - converts to sentence case (first letter capitalized)
 - Substring functions for extracting parts of strings by position
 - `trim()` - removes extra spaces from text

Alphabetical sort order for mixed text and numbers

- We alphabetize words starting with the first letter, then moving to the second when the first ties, etc.
 - Aardvark before Adam
- Shorter strings before longer that start with them
 - an before and
- Numbers come before letters
- If mixed, Excel considers numbers to be text characters, and sorts them alphabetically, not in numeric order
 - a1, a10, a2 instead of a1, a2, a10
 - What about 1a, 2a, 10a?
- To get the numbers to sort properly, use leading zeros (01a, 10a, 02a), or keep letters and numbers in separate columns

Dates and times

- 1/1/17, Jan1, and January 1, 2017 would all be recognized by Excel as 1/1/2017
- When Excel guesses you entered a date, it:
 - Converted the date internally to a number
 - Sets the display type to date
- Excel uses 1/0/1900 as the “epoch”
 - All dates stored as number of days since this date
 - Internally, 1/1/17 is 42736
- Be careful, because:
 - What is stored internally is not the same as what is entered
 - What is displayed is not the same as what is stored internally
 - What if Jan1 refers to the first measurement you made on Jan?

Time

- Time is recorded as a fraction of a day
- If the time recorded is 2:00 pm, this is converted as:
 - $14 + 0/60 = 14.00$ ← decimal minutes
 - $14.00/24.00 = 0.5833333333333333$ ← decimal hours and minutes
- If a date is given as well, it is the whole number in the date/time number
 - 2:00 pm on 1/1/17 is 42736.5833333333
 - Note: total of 15 decimal places, so storing a date means storing time to only 10 decimal places
 - Not a huge problem – 1×10^{-10} days is 0.00000864 seconds, or 0.0084 milliseconds (still precise time keeping)

Advantage of storing dates/times as numbers

- Can do math on them
 - Days between two dates, time elapsed between two times easy to calculate
 - Sequences of days easy to generate
- There are special functions that can be applied to date/time data
 - Extracting portions: month(date), day(date), year(date), hour(time), minute(time), second(time)
 - Calculate the weekday, week number of any date: weekday(date), weeknum(date)

Logical

- Logical data types can only take one of two values, TRUE or FALSE
- Used for making decisions – results of comparisons produce logicals
- Displayed as text (TRUE or FALSE), internally represented as numbers (1 or 0)
- Logicals are centered in the cell, not left-justified like text, or right-justified like numbers
- Can do math with them (but functions may not work properly with them)

Logicals in a spreadsheet

The screenshot shows a spreadsheet interface. The formula bar at the top contains the formula `=RANDBETWEEN(0,1)=1`. The spreadsheet grid shows a column of logical values: FALSE, FALSE, FALSE, FALSE, FALSE, FALSE, FALSE, FALSE, FALSE, FALSE, TRUE, TRUE, FALSE. A secondary window titled "Entering logicals" is open, showing a grid with columns D, E, and F. The values in this window are: true, FALSE, TRUE, TRUE, FALSE, FALSE.

| A | B | C | D | E |
|---|----------|---|----------|---|
| | Logicals | | Add zero | |
| | TRUE | | 1 | |
| | FALSE | | 0 | |
| | TRUE | | 1 | |
| | TRUE | | 1 | |
| | FALSE | | 0 | |
| | FALSE | | 0 | |
| | | | | |
| | Text | | | |
| | TRUE | | #VALUE! | |
| | FALSE | | #VALUE! | |
| | TRUE | | #VALUE! | |
| | TRUE | | #VALUE! | |
| | FALSE | | #VALUE! | |
| | FALSE | | #VALUE! | |

Can do math with them!