

This first homework is a warm up to thinking about functions. It is the only homework in Excel; all problems can be completed with standard Excel functions (formulas that begin with =). A list of Excel functions is found here: <http://www.excelfunctions.net/ExcelFunctions.html>.

Enter numbers/strings in green cells
Enter formulas (starting with "=") in blue cells. These formulas should update when new numbers/strings are entered in green cells.
Enter responses/text in red cells

Note that there are **two sheets**, one for each part. Write an estimate of the time it took to complete each part at the bottom of each spreadsheet.

Part 1: Using Excel to Manipulate Numbers and Strings

A *variable* (e.g., x , y , A , B , ...) is a symbol that represents a value. In the Excel spreadsheet, green cells *assign* a value to a variable. For example, the value 5 is *assigned* to the green cell B12; thus, when we refer to the variable B12 we mean the value 5. A *function* is a cell that takes numbers (e.g., 5) or references to other cells (e.g., B12) and computes a new value. In Excel, these functions will begin with an equals sign.

Manipulate numbers in Excel. Write functions in the blue cells that calculate the expressions in the first column using the variables x and y . When you enter different numbers in the green cells, the blue cells should automatically update:

variable x:	3
variable y:	-2
x+y=	1
x-y=	5
y-x=	-5
x*y=	-6
x/y=	-1.5
y/x=	-0.6666667

variable x:	5
variable y:	-2
x+y=	3
x-y=	7
y-x=	-7
x*y=	-10
x/y=	-2.5
y/x=	-0.4

What values of x and y “break” some of these functions? Which functions are “broken”? Write your answer in the spreadsheet.

Manipulate Strings in Excel. When an expression describes a string, variables represent strings. In the green cells, strings A and B can be words or phrases. The character c is a single letter. Write functions in the blue cells that calculate the following quantities.

- `length(A)` computes the number of characters in the string variable A .
- `A+B` “glues” the string variables A and B together. If $A=Hi$ and $B=There$, $A+B=HiThere$. What’s one difference between adding strings and adding numbers? Write your answer in the spreadsheet.
- `remove(A,c)` makes a string from A by removing all occurrences of the character c . If $A=she\ sells\ seashells$ and $c=e$, then `remove(A,c)=sh\ slls\ sashlls`.
- `occur(c,A)` computes the number of times character c appears in string A .

You might need to look carefully at the Excel Functions page and/or use outputs from previous functions). Changing the strings A and B and the character c should change the function values.

HW1.2: Using Excel to Calculate Hamming Distance

Calculate Hamming distance between two strings. The *Hamming distance* between two strings of equal length is the number of characters that are different between them.

CAT and CAT have a Hamming distance of 0.

CAT and BAT have a Hamming distance of 1.

CAT and BAG have a Hamming distance of 2.

In the spreadsheet strings A and B are seven characters long. Place one character in each of the green cells. Using Excel functions, compute the Hamming distance between strings A and B . You may need to use an intermediate function that fills in the blue cells at each position.

Calculate Hamming distance of an alignment. When we shift the top string to the left or the right, the relative positions of the strings are important. This is an *alignment* of two strings. A character matched with nothing has a distance of 0, and the length of the strings don't need to be the same.

MOUSE
GROUSE

Hamming Distance=5

MOUSE
GROUSE

Hamming Distance=1

MOUSE
GROUSE

Hamming Distance=4

Calculate the Hamming distance between the two shifted strings. You will need to handle the case where a letter is matched with a blank cell.

Minimizing Hamming distance. For this part, you will now shift the strings A and B to get the smallest Hamming distance possible.

1. In the green cells, write an alignment of A and B that produces the smallest Hamming distance possible.
2. There are multiple alignments that produce the smallest Hamming distance. Write another configuration.

Extra questions (Optional).

1. If you were going to automate this process, how many configurations would you have to check? Justify your answer in the spreadsheet.
2. I claim that the smallest Hamming distance (as defined above) between *any* two shifted strings in an alignment is at most one. Why?
3. Propose a new measure that accounts for the artifact above. In other words, write a definition where the smallest distance may be greater than one.