1 Making decisions

We use the `if` function to express decisions in Excel. Before using it, we write a sentence in English that describes our decision making process. The following sentence is one that might be used as a classification tool.

> If the distance from home to school is more than a mile then you are a driver else you are a walker.

We need to pull three things out of that sentence. The first is the *logical test*. The logical test is the part that specifies how to make the decision. In our case, it is "the distance from home to school is more than a mile". We convert this phrase to Excel syntax by noting that the distance from home to school is stored in column K in our Class Data spreadsheet. Since we always focus on just one row when developing a formula, let's pick row 2. So we capture the notion of "the distance from home to school" with the cell reference `K2`.

Logical tests include the use of some common operations (called *relational operations*, including less than (whose syntax is `<`), greater than (`>`), equal to (`=`), less than or equal to (`<=`), greater than or equal to (`>=`), and not equal to (`<>`).

In our case, the English description says, "is more than" which is equivalent to "greater than", so we get a logical test: `K2>1`. Question: Which relational operation is equivalent to "is at least"?

The second part to pull out of the English sentence is the action that Excel should carry out in the case that the logical test evaluates to true. The action is usually one of two things: display a word or sentence or evaluate a formula and display the result.

To display a word or sentence, all we need to do is put the word or sentence inside of two sets of double quotes: `"Driver"`. To display the result of a formula–any formula, including one that calls functions–just type in the formula without the usual leading equal sign (e.g. `D2-2`).

The third part to pull out of the English sentence is the action that Excel should carry out in the case that the logical test evaluates to false. This part is similar to the second part. Putting it all together, we get the following syntax: `=if(K2>1,"Driver","Walker")` to place in a cell in some column in row 2.

2 Pulling the English sentence out of a problem statement

Usually, we are not given the English sentence given in the previous section. We are much more likely to see a description like this:

> We are trying to figure out the maximum number of parking spots that we might need on campus. We have a collection of data that lists the distance each of our students is from campus. Let's assume that anyone who lives more than a mile from campus will regularly drive and that everyone else will walk. Can you tell me the names of students who are walkers and who are drivers?

First we need to recognize this problem as a classification problem–we are trying to put each of our students in one of any number (in this case two) of classes. Essentially, we need to make a *decision* about each of the students. Then we need to know that the IF function in Excel is the tool we use to make decisions of many kinds, but especially this kind.

The description has a lot of information, some of it is important to answering the question, some of it really is not. When we look at the description of the task, the place to begin is at the end. What are we being asked to do or to look for? What task have we been given? In this case, the question is,

> Can you tell me the names of students who are walkers and who are drivers?

The general form of the English sentence we are trying to write is:

> If *some decision that evaluates to either true or false*
> then *some action that takes place when the decision is true*
> else *some action that takes place when the decision is false*

After studying the question we easily come to the conclusion that "tell me the names of students who are walkers" and "tell me the names of students who are drivers" are pretty important to this problem. Notice that I "distributed" the "tell me the names of students" part over the other parts, much like we distribute multiplication over addition in algebra. Once we have these two pieces of information, we need to figure out which part of the If sentence they correspond to. Do the correspond to a decision or an action? In this case it is pretty clear "tell me" is an action. So it looks like we have two different actions–perfect for the If sentence. We are missing the decision. How do we figure that out? We go back to problem description, to the next to last sentence. they correspond to.

> Let's assume that anyone who lives more than a mile from campus will regularly drive
> and that everyone else will walk.

Buried in this sentence is the decision we are looking for. It tells us *how* to decide whether someone is a walker or a driver: "anyone who lives more than a mile from campus." We now have our three parts. Plugging them into the general form, we get:

> If *anyone who lives more than a mile from campus*
> then *tell me the names of students who are walkers*
> else *tell me the names of students who are drivers*

Of course, this isn't very pleasing English to read, so we need to clean it up a little bit:

> If *someone lives more than a mile from campus*
> then *that person is a walker*
> else *that person is a driver*

After a moments reflection, you may realize that this isn't right. In fact, we have to be very careful about deciding what is the *true action* and what is the *false action*. Furthermore, we cannot make that decision until *after* we have decided what the decision is going to be. Here's the final sentence:

If *someone lives more than a mile from campus*
then *that person is a driver*
else *that person is a walker*

What would the rest of the sentence be if it began, "If someone lives less than a mile from campus"? How would you translate that sentence into an IF function in Excel?

3 Categorizations that depend on two or more categories

Some problems depend on two or more categories. Consider this problem:

We want to know which of our students are upperclassmen and which are underclassmen. Remember in order to be an upperclassman someone must either be a junior or a senior.

Let's practice writing an English sentence. This is a good example to work on because it doesn't follow the pattern that we saw in the previous section. Nonetheless, we will start with the last sentence. Playing around with it, we see that "be an upper-classman" is an action and it looks like "someone must either be a junior or a senior" is a decision. Now we have a decision and an action. We get the other action from the first sentence, sort of. The first sentence mentions underclassmen and so we have to put things together to get a second action of "be an underclassman". Our English sentence becomes:

If *someone is a junior or a senior*
then *that person is an upperclassman*
else *that person is an underclassman*

The decision part of this sentence confronts us with a new type of decision where either of two possibilities can satisfy the decision. Thankfully, Excel has a built in tool called the OR function. Unfortunately, the OR function doesn't understand things the way we do. Remember in the previous section when I "distributed" one phrase over another phrase? Well, we need to do it again, but this time it is so that we can do a better job of converting our English sentence into an Excel function. Here's what we get:

If *someone is a junior or someone is a senior*
then *that person is an upperclassman*
else *that person is an underclassman*

Let's focus on the decision part: someone is a junior or someone is a senior. In my spreadsheet, column H holds the academic status. If I want to check whether someone is a junior, I would write `H2="junior"`, and to test whether someone is a senior, I would write `H2="senior"`. I use the OR function to put them both together to make my decision part: `OR(H2="junior",H2="senior")`. If we try to read that decision part it comes out something like: "Or someone is a junior, someone is a senior"–not very pleasing to the ear, certainly! Nonetheless, Excel demands that we write it that way. Essentially, Excel wants to know from the beginning what function it is using. It doesn't want to wait until the middle of the formula.

Now we can put the entire cell contents together as:
`=IF(OR(H2="junior",H2="senior"),"upperclassman","underclassman")`

Here's another type of problem that you might encounter.

> We have a special prize to award to a non-traditional student who will be graduating this year. We want to know all of the non-traditional students who are seniors. I guess any senior who is older than 25 is non-traditional.

Converting this problem into an English sentence requires some solid word sleuthing. As usual, we will start with the last sentence. It gives us the definition of non-traditional. The person must be a senior and older than 25. The other thing that is a bit unusual is that this problem asks us to identify only the non-traditional seniors, no one else. This suggests that the "else" part of our If sentence is not needed.

> If *someone is a senior and their age is more than 25*
> then *that person is an non-traditional senior*

In the decision part of this If sentence, we notice the word "and". Compare that to the decision part of the previous example. In that case we used the word "or". As was with the case with OR, Excel has an AND function. Before we get to using it, though, we need to figure out how to test both whether someone is over 25 years old and whether someone is a senior. We tested whether someone was a senior in the previous example (`H2="junior"`). In my spreadsheet, a person's age is recorded in column B, so to find out if a person is older than 25 I use `B2>25`. (Why is `B2>=25` incorrect? Why is `B2>"25"` incorrect?)

Putting these expressions into the AND function is a exactly like putting expressions into an OR function: `AND(B2>25,H2="senior")`. Again, reading this out loud is not very pleasing, "And the age is greater than 25, the academic status is senior." Putting these pieces together, the cell entry is: `=IF(AND(B2>25,H2="senior"),"non-traditional senior","")`

Notice that the false action is two double quotes right next to each other. The syntax rules for the IF function don't require a false action, but if you don't put one in, the word FALSE will appear in cells where the decision evaluates to false. The juxtaposed double quotes refer to the empty string and quite literally stand for the concept of nothing and directs Excel to place nothing in the cell whenever the decision is false.

4 AND, OR: The rules

If you study the syntax of the OR function, you are allowed to put two or more expressions. If any one of those expressions evaluates to true, then the OR function evaluates to TRUE. The only other possibility is for an expression to evaluate to FALSE. In order for the OR function to evaluate to FALSE, every one of the expressions must evaluate to FALSE.

The AND function has a similar syntax rule and description. If any of the expressions evaluates to FALSE, then the AND function evaluates to FALSE. In order for the AND function to evaluate to TRUE, every one of the expressions must evaluate to TRUE.