Excel Lesson: **Using Excel to Determine if a Value is Within any Group of Contiguous Values**

Topics Covered:

Worksheet Functions:

MATCH()

INDEX()

Real World Example Application:

Given a list of things with a value that falls into one of several value range groups, determine which group it falls within. More: You have a list of people with their ages, you need to determine which of several age groups they fall in. The age groups are, of course, based on current age and represent a 'contiguous value range', i.e. each range picks up where the last left off.

Discussion: In essence you're trying to determine if a value (individual's age) falls within a group, or range, that is defined with an upper and lower level. For example, the group INFANTS might include all ages from 0 to 2 years old. Each group's range must be exclusive and it must pick up where the previous one left off for this to work properly. The next group in this list might be SMALL CHILDREN and you might see it as 2 to 6. You must make a determination if a child 2 years old is an INFANT or a SMALL CHILD, so you have to really have that "2" defined. For argument's sake lets say we, or someone else, has decided that a child over 1 year old but less than 6 years old is a SMALL CHILD, so the first range really becomes: 0 to up-to-but-not-including-2 years old. And we go on to find that the age 6 is really the start of the next age group (that is, SMALL CHILD includes all that are at least 2 years old, but are UNDER 6 years old, which could be rewritten as 2-5).

We finally find that our age groups are:

Minimum Included Age (yrs)	Maximum Included Age (Yrs)	Name of Age Group
0	1	INFANT
2	5	SMALL CHILD
6	12	CHILD
13	19	TEEN
20	29	YOUNG ADULT
30	65	ADULT
66	No limit	OLDER ADULT

We now have all the information we need to set up a pair of data sets on an Excel worksheet that we can use to determine which group, by name, which an individual falls into based upon their age.

They will be two separate data sets, but they are related in that the order that each one's list gets put into must be the same.

On any sheet in the workbook, set up the following list for the ages. It is not a mistake that I have not included the ages from the Minimum Included Age column – it is not needed, and it can be inferred or derived from the list of Maximum Included Ages. And Excel doesn't need it to figure things out! Be careful, because you'll notice that I've entered the ages in reverse, or descending order. This is to suit the needs of Excel's MATCH() function later on.

I'll set my list up in column G, beginning with row 1 and putting a heading there although the heading is not required. But it will help others understand what the information represents later.

G
Max Age of Group
200
65
29
19
12
5
1

For convenience sake, and to help anyone looking at the lists later on, we will set up the corresponding list of named age groups right next to this one on the same sheet, in column H like this:

H
Named Age Group
OLDER ADULT
ADULT
YOUNG ADULT
TEEN
CHILD
SMALL CHILD
INFANT

Next we will take a look at the 'raw' information we have to work from and set things up to figure out which group each individual is in.

Let's assume that our list is fairly simple: a name and a current age on the same sheet in columns A, B, C and D with the two data sets we've set up in columns G and H. It looks like this:

NAME	Age	Group#	Group
Mabel	27		
Arthur	13		
Mary	6		
Tim	1		
Sam	72		
Ralph	65		
Joe	8		
Helen	37		
Bea	42		
Harold	18		
Cindy	19		

We're going to take this one step at a time so that you will not have to swallow too big a bite to begin with. Before beginning, a little discussion of the two functions we will be using needs to occur.

MATCH() : The MATCH() worksheet function needs 3 pieces of information –
1 – What to match (our individual's age – in column B beginning at row 2).
2 – Where the list to match with is located (it is in *range* $G$2:$G$8)
3 – How to match.

If you're confused about the $G$2:$G$8 entry, the $ keeps the value behind it from changing – our list is always going to be in that range and we don't want it changing if we copy or extend any cell formula that contains reference to it.

How to match. This can have one of 3 values –
1 (the default value if you leave it out completely) If match_type is 1, MATCH finds the largest value that is less than or equal to lookup_value. Lookup_array must be placed in ascending order: ...-2, -1, 0, 1, 2, ..., A-Z, FALSE, TRUE.

0 (zero) If match_type is 0, MATCH finds the first value that is exactly equal to lookup_value. Lookup_array can be in any order. **NOTE:** We cannot use this one because it looks for that ***exact*** match, and most of the time we won't have one!

-1 (negative one) If match_type is -1, MATCH finds the smallest value that is greater than or equal to lookup_value. Lookup_array must be placed in descending order: TRUE, FALSE, Z-A, ...2, 1, 0, -1, -2, ..., and so on. **NOTE:** This is the one we will use because it fits the way we've set up our upper limits for age groups in column G. We could have used the first one if we had set it up differently, based on lower age limit of the groups. But we didn't, so here we are.

MATCH() does not return the actual value in a cell in the list. It returns the position in the list of the item that met the match criteria. So our return values will be from 1 (OLDER ADULT) to 7 (INFANT). If no match is found you'll get a #N/A error indication in the cell that uses the MATCH() function.

Let's set up a formula to find out just where in the list our match occurs. Type the following into cell C2 right next to Mabel's age of 27 (you can type in lower case if you like, Excel will fix it all up later – I'll show what it looks like when it is done). Press the [Enter] key to complete the typing.

=MATCH(B2,$G$2:$G$8,-1)

You should see a **3** in cell C2 at this time. This indicates that the match was found at the 3rd entry in the list (29) which is correct. At 27 years old, Mabel is in the YOUNG ADULT group (ages 20 through 29, inclusive).

Well, that doesn't do us much good – we need the words YOUNG ADULT, not some number that we now have to use to look something else up somewhere else. We're not much better off than when we started. Have faith – it will all straighten itself out in the end.

Copy that formula, or drag (extend) it down through all the cells next to an age in column C. You'll see various numbers pop up, indicating which group number held the match for that individual's age.

NAME	Age	Group#	Group
Mabel	27	3	
Arthur	13	4	
Mary	6	5	
Tim	1	7	
Sam	72	1	
Ralph	65	2	
Joe	8	5	
Helen	37	2	
Bea	42	2	
Harold	18	4	
Cindy	19	4	

Think about what you have just done: You have created an ***index*** into the age grouping data set that tells which position in that list the match was found at. Excel also has a worksheet function, **INDEX()** that can use a value as an index into a list to return the actual contents of the cell at that position!!

If you look at Excel Help for INDEX you'll find it has two forms, an *array* syntax and a *reference* syntax. We will be using the *array* syntax since it returns the value in the cell it locates.

Remember that cells in column C at C2 through C12 now hold an index value for us, courtesy of the MATCH() function formula in those cells.

The **INDEX()** function (in the array syntax we are using) needs 2 pieces of information:
First it needs a reference to the location of the list to find something in, and
Second it needs the index of the item in the list. 1 = the first entry.

Lets put the following formula into Cell D2 (out from Mabel's information again) –
=INDEX($H$2:$H$8,C2)

The reference to $H$2:$H$8 is a reference to our list of age group names. As before, we don't want that changing as we copy/extend our formula down the sheet. The C2 part says we want to use the value in cell C2 as the INDEX. If you typed it right, you should see this in your table now:

NAME	Age	Group#	Group
Mabel	27	3	YOUNG ADULT
Arthur	13	4	
Mary	6	5	
Tim	1	7	
Sam	72	1	
Ralph	65	2	
Joe	8	5	
Helen	37	2	
Bea	42	2	
Harold	18	4	
Cindy	19	4	

And if you copy/extend the formula in cell D2 on down through C12, then it should look like this:

NAME	Age	Group#	Group
Mabel	27	3	YOUNG ADULT
Arthur	13	4	TEEN
Mary	6	5	CHILD
Tim	1	7	INFANT
Sam	72	1	OLDER ADULT
Ralph	65	2	ADULT
Joe	8	5	CHILD
Helen	37	2	ADULT
Bea	42	2	ADULT
Harold	18	4	TEEN
Cindy	19	4	TEEN

Voila! You've now figured out which group everyone belongs in, and in effect you have taken each individual's age and determined what range (min-max) they fell into.

But doesn't it seem a bit clumsy? You have an extra column (Group#) that you really aren't interested in – and it just takes up extra space on the sheet and might even confuse people.

Simplifying Things By Making Them More Complex
Boy, if that doesn't sound like it came right out of a Military Intelligence handbook for oxymorons. But we can.

Let's revisit the formula used in column D (the INDEX() formula). In cell D2 it looks like this:
=INDEX($H$2:$H$8,C2)

We notice that we have a reference to cell C2, which as we remember actually contains another formula that gave us a value that we are using in this one. What if we replace the "C2" entry in this formula with the actual formula we now have in cell C2? Let's try it. Change the formula in cell D2 to look like this:
=INDEX($H$2:$H$8, MATCH(B2,$G$2:$G$8,-1))

You should still see YOUNG ADULT in the cell when you're done! If you don't then recheck your formula – make sure you have the right number of (and) in it and that they are in the right places. Make sure you didn't stick an extra = sign in the middle of it. Those would be the two most likely mistakes to make at this point, I would think.

Copy/extend that formula down through Cindy's entry on the sheet and things should all still look just the same.

Want to prove that it still works right? Highlight everything in column C (Group#) and delete it! Or even delete that entire column – it should all still look correct:

NAME	Age	Group
Mabel	27	YOUNG ADULT
Arthur	13	TEEN
Mary	6	CHILD
Tim	1	INFANT
Sam	72	OLDER ADULT
Ralph	65	ADULT
Joe	8	CHILD
Helen	37	ADULT
Bea	42	ADULT
Harold	18	TEEN
Cindy	19	TEEN

Max Age of Group	Named Age Group
200	OLDER ADULT
65	ADULT
29	YOUNG ADULT
19	TEEN
12	CHILD
5	SMALL CHILD
1	INFANT

If you're using Excel 2000 or XP, it even took care of 'fixing' the references to columns G and H by changing them to F and G in the formulas in the new column C. Some

Earlier versions of Excel may not do that properly and you'll have to fix the first formula and re-copy or extend it to get things looking right again.

When will this fail? If you find someone older than 200 years old. I chose that for an upper limit for the last group because it seems an extremely unlikely age to encounter in the average population. Want an impossible number? Try 1000 – works just as well, and I don't think we've had anyone of an age even close to that since Biblical times.

Couldn't I accomplish this with a bunch of IF() statements? In this case, yes you could – but if you think my last formula looked complex, you are in for a surprise! The equivalent nested IF() statement would look like this:
=IF(B2>65,"OLDER ADULT",IF(B2>29,"ADULT",IF(B2>19,"YOUNG ADULT",IF(B2>12,"TEEN",IF(B2>5,"CHILD",IF(B2>1,"SMALL CHILD","INFANT"))))))
And that is a lot more difficult to deal with if the age group range limits change in the future!
Also, there is a limit on the number of nested IF() statements. You can only nest IF() statements 7 deep. We have 6 in the above formula. Add two more group categories and you cannot use nested IF() at all. The method shown in this paper has no practical limit (actually, it's limited to a list of 65,535 items, which is the maximum number of rows allowed in an Excel XP worksheet – some earlier versions have lower limits).

Making it Easier to Maintain:
One way to make this easier to maintain, if you anticipate adding more group categories or perhaps moving the data lists to another sheet in the workbook (it will still work just fine, just have to reference the address of the lists differently), you should consider giving each of the two lists a NAME reference. It is beyond the scope of this paper to discuss how to create and use NAMED references, but it is an incredibly worthwhile subject to check up on – it can make workbook maintenance and formula creation much simpler and reduce the chance for unexpected results later on.

Do a search in Microsoft Excel Help for "**Define named cell references or ranges**" for more information on how to work with cell/range names; how to give them names and how to use the names later on.

Wouldn't it have been nice to refer to **AgeRangeList** instead of $G$2:$G$8 and perhaps **AgeGroupNames** instead of $H$2:$H$8 in the formulas we created earlier? Using NAMED Cells and NAMED Ranges makes that possible – and Excel will keep up with things as you add to/delete from the lists that are NAMED Ranges, so you never have to change a formula as your lists expand and contract.

ENJOY