Microsoft Excel: MLB Payrolls exercise

This is a beginner exercise for learning some of the most commonly used formulas and functions in Excel. It uses an Excel spreadsheet called "MLB Payrolls 2009_2011" that has two worksheets. We're going to use the worksheet called "MLB Payrolls," which has one record for each team in Major League Baseball, with three columns of numbers. These columns represent the total each team spent – each year – on player salaries.

Our goal, as journalists, is to look for some patterns and trends in this information.

Let's start by thinking of some questions we might want to know about the money spent on player salaries in professional baseball.

Some possibilities:

Which team had the greatest increase in its spending?
Which team had the greatest decrease?
How much money was spent overall in the league on salaries and how has that changed over time?
What's the average payroll in the league?
Who was the highest paid player?
What's the average player salary?

Can we answer all of these questions with this particular data? The answer is no. This worksheet only shows the totals spent by each team; we don't have any details on specific players. So we can't see the highest paid player and we also can't calculate the average player salary. To get those we would need data that shows one record for each player, including their specific salary for a given year. (you can see a slice of that in the other worksheet called "Twins-Brewers"; this contains the individual player salaries just for the players on those teams)

Now, we need to translate those questions (that we can answer) into language that Excel will understand. All of these questions require math, which is what Excel was made for.

Let's start with "which team had the greatest increase in spending?"

This can be a tricky one. Of course, we can take the 2011 total and subtract the 2009 total to see which team had the largest increase in raw dollars. This would show us that the Phillies spent about $59 million more in 2011 than they did in 2009. But is this a fair way to compare teams?

Look at the Kansas City Royals. In 2011 they only spent $36 million. The Phillies on the other hand spent $172 million. And the Yankees spent over $200 million. The difference that the Phillies spent in 2011 compared to 2009 is bigger than the total that the Royals spent in 2011.

This is a situation where we are trying to compare entities of different sizes and we need to put them on a level playing field first. Think of it this way: the Phillies with their huge salary budget probably see a $59 million increase as relatively small. The Royals would probably view that as an enormous change.

Instead of using the raw dollar change, we can use a percentage change – that will tell us the relative size of the change in comparison to the others.

To do that we're going to use a formula in Excel. All formulas start with an equal sign and use the cell addresses to do the work (notice that the worksheet contains letters going across the top and numbers going down the side. Click on a cell and it will highlight the letter and number that make up its address. For example, in MLB Payrolls, the header that says "team" is located in cell "A1")

A percentage change is calculated by taking the new number minus the old number and dividing by the old number.

=(N-O)/O   (think of "No!" as a way to remember this)

For this formula we will take D2 (the 2011 number) minus B2 (the 2009 number) and divide by B2. You type the formula in cell E2, which should be the first row of data below the headers. (Yours might have a different team at the top than the picture below)

From Access	From Web	From Text	From Other Sources	Existing Connections	Refresh All	Edit Links		Sort	Filter	Advanced	
		Get External Data				Connections			Sort & Filter		

SUM   ✕ ✓ fx   =(D2-B2)/B2

	A	B	C	D	E	F
1	TEAM	2009	2010	2011	Pct Chg	
2	Philadelphia Phillies	$113,004,046	$141,928,379	$172,976,379	=(D2-B2)/B2	
3	Minnesota Twins	$65,299,266	$97,559,166	$112,737,000		
4	Boston Red Sox	$121,745,999	$162,447,333	$161,762,475	⊕	
5	San Francisco Giants	$83,026,450	$98,641,333	$118,198,333		
6	Chicago White Sox	$96,068,500	$105,530,000	$127,789,000		
7	Los Angeles Angels	$113,709,000	$104,963,866	$138,543,166		
8	Texas Rangers	$68,178,798	$55,250,544	$92,299,264		

After typing, hit Enter. Then get the copy tool (a thin black cross) in the lower right corner of the cell that has this new answer and drag it down to the other rows.

You may need to highlight the column, then right-mouse click and choose "Format Cells" to change the formatting to percentage.

	E	F
	Pct Chg	
379	53.07%	
000		
175		
133		

Once you've copied it down, put your cursor somewhere inside the chunk of data (doesn't matter where, just make sure it's not in an empty cell) and then go to the Data menu and choose "Sort"
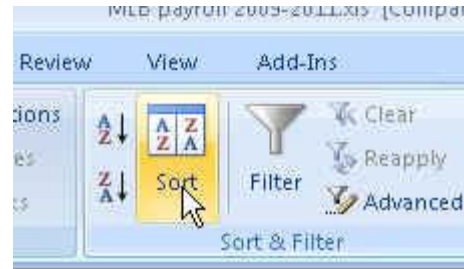
That will bring up a dialog box where you can choose which field to sort by. Notice that the box in the upper right corner is checked for "My data has headers" and then the pull-down menu on the left (next to "Sort by") lists your headers. Choose the one you just created for "Pct Chg". When you do that, the pull-down menu on the right will change to "Smallest to Largest". You can choose to change that to "largest to smallest" so that we see the team with the biggest increase at the top of our list.
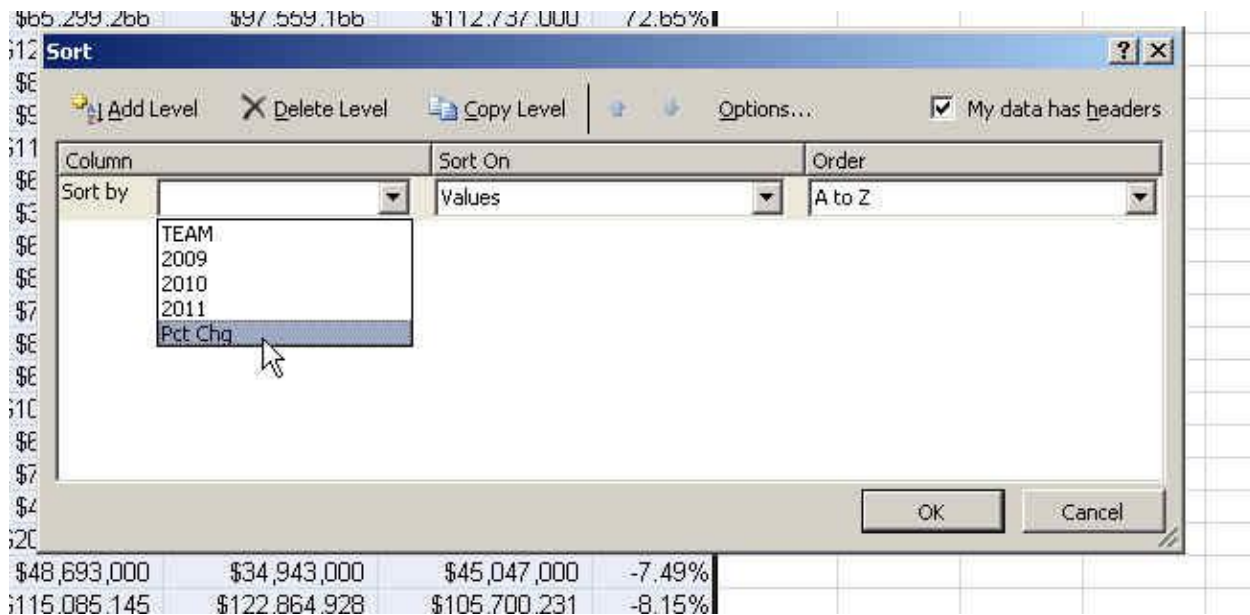
Hit OK and your list should now sort with Minnesota Twins at the top – a nearly 73% increase. The Royals are at the bottom with nearly 50 percent decrease.

The next question we wanted to ask was "how much has been spent on salaries in the league overall, each year?" To do that we need to add up all the payrolls for each year.

Excel has a wonderful function called SUM that allows us to add up all the values that are in a range of cells. In this case, the 2009 payrolls are in cells B2 through B31. The 2010 payrolls are in the range from C2 to C31, etc.

We'll put these totals on a row below our table. So go to cell A33 (leaving an empty row between our data table and these new totals) and put the word "Totals" in cell A33. This will be our label to remind us what this is.

Then in cell B33, we'll put our formula:

27	Arizona Diamondbacks	$73,516,666	$60,718,1
28	Houston Astros	$102,996,414	$92,355,5
29	Tampa Bay Rays	$63,313,034	$71,923,4
30	Cleveland Indians	$81,579,166	$61,203,9
31	Kansas City Royals	$70,519,333	$71,405,2
32			
33	TOTALS	=sum(b2:b31)	
34			
35			
36			

Hit enter and then get the copy tool and drag the formula across to columns C and D. It's possible your answers might look like this:

29	Tampa Bay Rays	$63,313,034	$71,923,471	$41,053,571	-35.16%
30	Cleveland Indians	$81,579,166	$61,203,966	$49,190,566	-39.70%
31	Kansas City Royals	$70,519,333	$71,405,210	$36,126,000	-48.77%
32					
33	TOTALS	$2,666,728,494	#############	############	
34					
35					

That just means that the columns are not wide enough to display the entire number. To make them wider, go up to the letters at the top. You can either drag the line between the C and the D or double-click on it to get it wider. Then do the same for the line between the D and E columns.

To answer the next question --- how has this changed over time – this time we can either calculate a percentage change or we can use a raw number (we're not comparing entities of differing sizes).

An easy way to get the percentage change is to drag the percentage change formula down from cell E31 so that it lands in cell E33. You'll end up with a Divide-By-Zero error message on the blank row (which you can just delete)

$70,694,000	-31.36%
$41,053,571	-35.16%
$49,190,566	-39.70%
$36,126,000	-48.77%
	#DIV/0!
$2,786,161,291	4.48%

So this 4% increase is simply telling us that the overall money
spent on player salaries in MLB is up 4 percent from two years' earlier. It's NOT exactly an average of the changes that we saw for each team (at least not in specific mathematical terms) but you'll notice that it's about smack in the middle of all the percent changes we calculated for the individual teams.

Next question: What's the average payroll?

Again, Excel has a function for this (two, actually). When you're dealing with dollar amounts – especially pay or housing values or things like that – you will want to look at both the AVERAGE and the MEDIAN. The average is calculated by adding all the values together and dividing by the number of records. The Median is the value that would be in the middle if you sorted all the payrolls and ordered them smallest to largest. An average can be heavily skewed by one or two "outliers" – numbers that are significantly bigger or smaller than all the other values. A median will not be skewed. When there are outliers, the median will more likely represent the "typical" value.

The rule of thumb is to calculate both and if they are close to each other, it's safe to use either one. But if the average is significantly higher or lower than the median, it's best to use the median (if you're attempting to tell your readers/viewers what the "typical" is).

To calculate these in Excel, we're going to use functions just like we did when summing our data.

Start a new line below the totals and put "Averages" in the A34 cell and "Medians" in the A35 cell. Then put these formulas in the B34 and B35 cell, respectively. Then copy them across to the other two columns of payrolls.

=AVERAGE(B2:B31)

=MEDIAN(B2:B31)

30	Cleveland Indians	$61,579,166	$61,203,966	$49,190,566	-39.70%
31	Kansas City Royals	$70,519,333	$71,405,210	$36,126,000	-48.77%
32					
33	TOTALS	$2,666,728,494	$2,717,759,865	$2,786,161,291	4.48%
34	AVERAGES	$88,890,950	$90,591,996	$92,872,043	
35	MEDIANS	$81,058,733	$84,325,333	$86,763,646	
36					
37					
38					

In this case, the averages and medians aren't too far off each other but you can see that the averages are higher – most likely due to the New York Yankees (the big outlier in baseball)

Let's do one last thing with this data. Let's figure out just how much bigger the Yankees are than the other teams. There are a couple things we can do.

The first is a "percent of total." We can calculate what percentage of the total MLB payroll (what we calculated on row 33) each team accounts for.

To do this, let's go to column F and put a header – "Pct Total" and then start a new formula in cell F2.

A percent of total is calculated by taking the team number and dividing it by the grand total.

After typing in this formula, hit enter and copy it down for the other teams.

	D	E	F	G
	2011	Pct Chg	Pct Total	
6	$112,737,000	72.65%	=d2/d33	
9	$56,944,000	54.60%		
9	$172,976,379	53.07%		
3	$118,198,333	42.36%		
4	$92,299,264	35.38%		
0	$127,789,000	33.02%		

Uh oh! We've got a problem.

	E	F	G
	Pct Chg	Pct Total	
7,000	72.65%	0.040463	
4,000	54.60%	0.613145	
6,379	53.07%	1.99365	
8,333	42.36%	#DIV/0!	
9,264	35.38%	#DIV/0!	
9,000	33.02%	#DIV/0!	
2,475	32.87%	#DIV/0!	
4,038	27.13%	#DIV/0!	
3,166	21.84%	#DIV/0!	
3,572	19.10%	#DIV/0!	
8,071	17.22%	#DIV/0!	
6,500	6.78%	#DIV/0!	

We can trouble-shoot by clicking on the answer in F3. If you double-click on it, it will highlight the cells that it is using to do the math and it will display the formula is the formula bar just above your data.

	D	E	F	G
	2011	Pct Chg	Pct Total	
,166	$112,737,000	72.65%	0.040463	
,719	$56,944,000	54.60%	=D3/D34	
,379	$172,976,379	53.07%	1.99365	
,333	$118,198,333	42.36%	#DIV/0!	
,544	$92,299,264	35.38%	0.455374	
,000	$127,789,000	33.02%	#DIV/0!	
,333	$161,762,475	32.87%	#DIV/0!	
,500	$95,304,038	27.13%	#DIV/0!	

What's wrong? We can see that the first cell address in the formula is correct. We want D3 (the payroll for the Marlins), but what's in D34?

It's grabbing the average we calculated for 2011 – not the total. If you go down to the ones that are showing a divide-by-zero error, you'll see that it's trying to get data from a cell that is empty.

We need to adjust our formula so that the divisor (the second cell address) is ALWAYS on the total that is located in D33.

Excel allows us to do that with what's called an "anchor"

To anchor a cell, we put a dollar sign ($) in front of the letter and/or the number of the cell address. To lock an address completely (regardless of whether you copy your formula down or to the right) you would use two anchors, like this:

=D2/D33

So go back to the Twins row – line 2 – and edit your formula (easiest way is to go up to the formula bar and edit there), adding in two dollar signs. Then copy the formula down again. This time you should have answers! (You may need to highlight the column, right-mouse click and choose "Format Cells" to change it to a percentage)

So this shows us that the Yankees account for 7% of the overall payroll spending in MLB.

Another way we could look at this is with a ratio. For every $1 spent by a team, how much do the Yankees spend?

To make this a little easier, I'm going to start by copying the Yankees information to a new line, below our data. (note: I didn't move it; just made a copy of it)

28	Houston Astros	$102,998,414	$92,355,500	$70,694,000	-31.36%
29	Tampa Bay Rays	$63,313,034	$71,923,471	$41,053,571	-35.16%
30	Cleveland Indians	$81,579,166	$61,203,966	$49,190,566	-39.70%
31	Kansas City Royals	70,519,333	$71,405,210	$36,126,000	-48.77%
32					
33	TOTALS	$2,666,728,494	$2,717,759,865	$2,786,161,291	4.48%
34	AVERAGES	$88,890,950	$90,591,996	$92,872,043	
35	MEDIANS	$81,058,733	$84,325,333	$86,763,646	
36					
37	New York Yankees	$201,449,189	$206,333,389	$202,689,028	
38					
39					
40					
41					
42					

Then start a new column – in G – and call it "Ratio"

Then in G2, we'll type a formula. A ratio is simply one number divided by the other. Whichever you want to be your "$1" team needs to go second. So we'll start with the Yankees (D37) and divide by the Twins (D2). Since we always want the first cell in our formula to be D37 (Yankees), then we also need to use anchors on that one.

Sort & Filter

	E	F	G	H
	Pct Chg	Pct Total	Ratio	
	72.65%	4%	=d37/d2	
	54.60%	2%		
	53.07%	6%		
	42.36%	4%		

So then if you format the cells to Currency and sort the table by this new column, you'll see that it has "$5.61" for the Kansas City Royals. That means the Yankees spend $5.61 for every $1 spent by the Royals. That puts these numbers down into a realm that every reader can understand.

More information on these formula can be found in the handout called "Easy Calculations in Excel"

Created by:
MaryJo Webster
Digital First Media
mwebster@digitalfirstmedia
updated March 2014