Lecture Note #4: Basic Functions

BUSI 201: Business Data Analysis

Spring 2024

Topic 1. The SUM Function

The sum function returns the sum of all cells included in the argument. So essentially, it is the + operator made simple. The syntax of the SUM function would be:

```
= SUM(ADDRESS OF CELLS CONTAINING THE VALUES YOU WOULD LIKE TO ADD)
```

Let us go through an example to make it clear. The cells C14 and E14 from this worksheet is supposed to track the sum of the quantity of goods, and the sum of the value of the goods in storage, so we may apply the SUM function.

Figure 1: SUM

The correct formula to find the total number of individual items in storage, which is the intended content of cell C14 is:

$$= \text{SUM(C4:C13)}$$

You could get the same result by manually adding all cells using the operator + as well:

= C4 + C5 + C6 + C7 + C8 + C9 + C10 + C11 + C12 + C13

Hopefully we can all agree that using the SUM function, even when there are only 10 entries, is by far the superior way to perform addition.
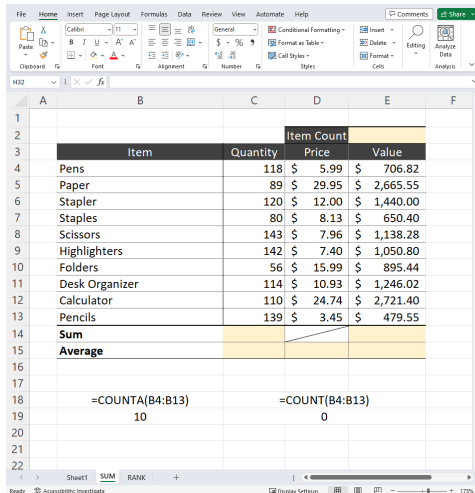
As the function performs an addition, only numerical information is allowed. Any values that are not numerical (including empty cells) will be ignored in calculating the sum. While this is probably not a limiting factor for most, the maximum number of arguments that you can input in the SUM function is 255. Please use the SUM function to fill out cell E14 to display the sum of the value of all office supplies in storage.

Topic 2. The COUNTA and COUNT Functions

The COUNTA Function

The COUNTA function will count the number of non-empty cells in the given range dictated by the arguments of the function. The syntax of the COUNTA function is identical to the SUM function.
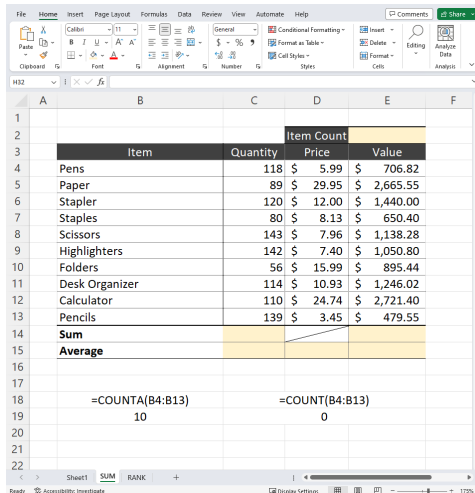

Figure 2: COUNTA

To report the number of unique items stored in storage for this firm, we can use the COUNTA function in cell E2. One possible forumla may be:

$$= \text{COUNTA}(B4:B13)$$

Note that it is stated that it is one possible formula. In fact, we can use any of the columns ranging from B to E to get the same answer in this specific case. This is possible as the other columns also have a value assigned to each row, matching that of the item description in column B. There is no one correct rule to use in real life, and the argument that you choose should depend on each specific case.

The COUNT Function

One commonly made mistake is using the COUNT function when you mean to use the COUNTA function. The COUNT function is a similar function, but only counts the number of cells in the range that contains numerical values.

Figure 3: COUNT vs COUNTA

Recall the status bar trick, where there were two separate types of "count" options. One was a "Numerical Count," and the other was simply "Count." In functions, for some reason, Excel decided to let COUNT correspond to the "Numerical Count," and COUNTA correspond to "Count."

The results depicted below the main table in figure 3 illustrates the difference in results when applying COUNT and COUNTA functions on cells with text data. Please try out applying the COUNT and COUNTA functions on the cell range C4:C13 to verify that they will return the same result, as column C consists of numerical data.

Topic 3. The `AVERAGE` Function

Figure 4: `AVERAGE`

The cells C15, D15, and E15 each aim to find the average value of the respective columns. The syntax is identical to the `SUM`, `COUNTA`, and `COUNT` functions. That is:

$$= \texttt{AVERAGE(C4:C13)}$$

Similar to the `SUM` function, all values related to this function must be numerical in nature. Any values that are not numerical (or empty) will be ignored in calculating the average value.

Please fill out the cells D15 and E15 and find the average unit price of each type of good in storage, and the average of each items' total worth, respectively.

Since the `AVERAGE` function returns the arithmetic mean of the selected values, the output from the `AVERAGE` function is identical to the result of taking the ratio of the `SUM` and `COUNTA` functions:

$$\texttt{AVERAGE(C4:C13)} = \frac{\texttt{SUM(C4:C13)}}{\texttt{COUNTA(C4:C13)}}$$

Topic 4. Other Statistical Functions

The `MEDIAN` and `MODE` functions share the same syntax with the `AVERAGE` function, and return the sample median and mode respectively. The sample median is the middle value of the sample data, while the mode is the value that most often appears in the set of data. You should learn more about these statistics in BUSI 205.

For instance, for a sequence of numbers $\{1, 2, 3, 5, 5\}$:

- `AVERAGE` $= 3.2$

- `MEDIAN` $= 3$

- `MODE` $= 5$

It is interesting to note that while the `AVERAGE` function "calculates" the value for us, the `MEDIAN` and `MODE` functions "locate" a certain value out of an array of numbers. Next we will be examining some more functions that "locate" certain values for the user.

3

Topic 5. Ordering Numbers: The `MAX`, `MIN`, `LARGE`, and `SMALL` Functions

The `MAX` and `MIN` Functions

Recall the other two variables that was readily available for the user in the status bar of Excel; the maximum and minimum values of an array of numbers. These are also available in the functions `MAX` and `MIN`, respectively. The syntax is identical to all previous examples, where the only argument(s) is the address of the cells that contain the numerical data you wish to evaluate. Please navigate to the `ORDER` sheet of the workbook `BUSI201-LEC04-Workbook.xlsx`.

Figure 5: Restaurant Menu Ratings

The worksheet consists of feedback from customers from an imaginary Italian restaurant. Suppose that we are primarily interested in what is the absolutely highest and lowest score any dish received. Those are the items to fill out in cells D2 and D3, repsectively. Each value can be found by using the `MAX` and `MIN` functions:

$$= \texttt{MAX(D7:D106)}$$
$$= \texttt{MIN(D7:D106)}$$

Similar to the functions we covered up to this point, the `MAX` and `MIN` functions ignore any cell that include non-numerical inputs and empty inputs.

The `LARGE` and `SMALL` Functions

Even with the `MAX` and `MIN` functions, we run into a problem with the next empty cell of D4. How do we find the second highest score in the list? Here is where the `LARGE` function comes into play.

Figure 6: `LARGE`

The `LARGE` function searches the array of numbers it is given, and looks for the kth largest value. Naturally, the `LARGE` function has two mandatory arguments; the address for the array of numbers, and the position of interest k. The function can be used in cell D4 as follows:

$$= \texttt{LARGE(D7:D106,k)}$$

In words, this means "Find the kth largest numerical value in cells D7:D106." Meanwhile, the `SMALL` function does the exact opposite, and returns the kth smallest numerical value in the given array of numbers. Note that:

$$\texttt{LARGE(ARRAY,1)} = \texttt{MAX(ARRAY)}$$

4

Topic 6. The RANK Function

We now know how to find the top or bottom ranked item in a list. But what if we wanted to know the ranking of all elements in a list? This is where the RANK function can be helpful. Navigate to the RANK worksheet which is a mock gradebook for a college level course.

Figure 7: RANK

The RANK function finds the relative rank of some number in a range of references. The syntax will be:

$$= \mathtt{RANK(number, ref, [order])}$$

You can read this function as "Find the rank of number out of a list of numbers in ref." Here, you can see that the last argument of the function [order] is in brackets. This means that this item is optional when writing up a function. If you want the greatest number to be ranked 1, the order should be 0. So, in our example, cell J4 should read:

$$= \mathtt{RANK(I4, \$I\$4:\$I\$33, 0)}$$