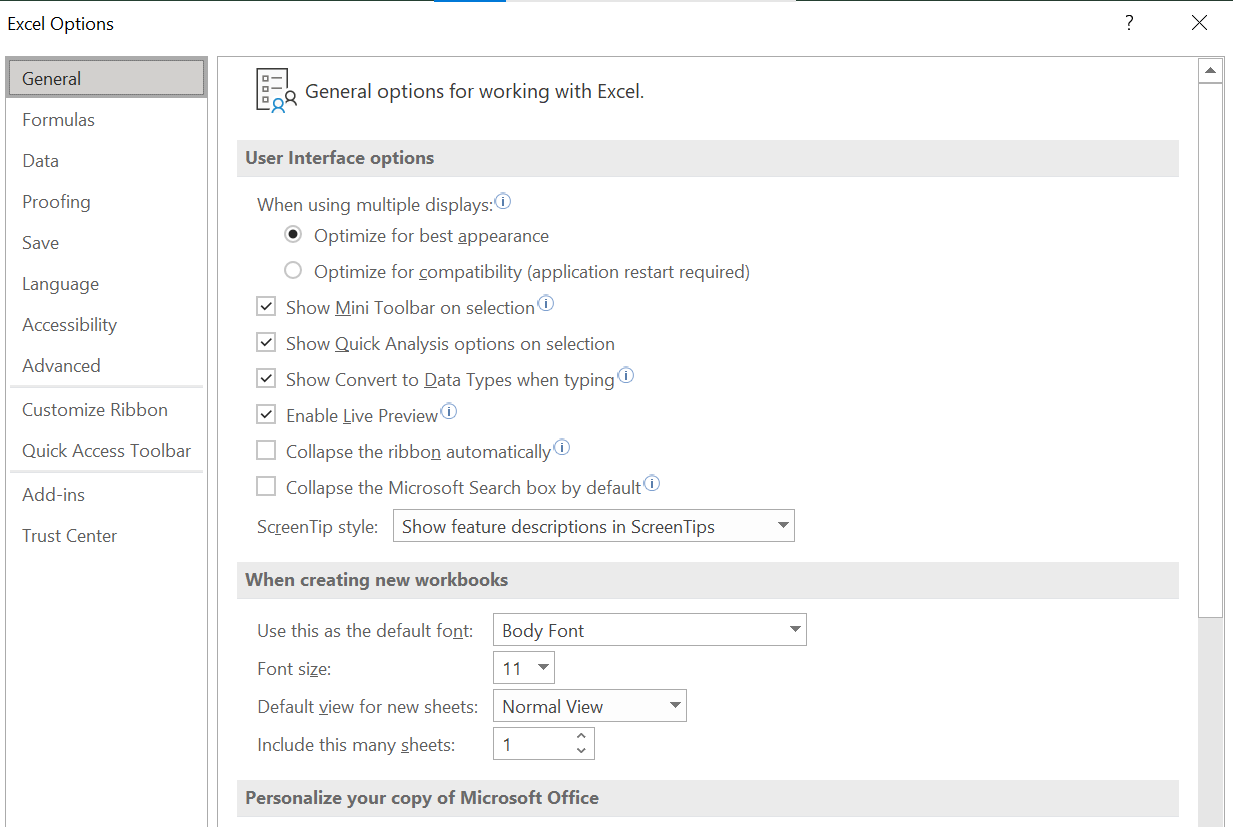
### **Introduction**

This guide walks you through how to analyse your data in Excel and use it to produce a results graph from your trial. Use it with the Excel version of the guide. You can also find a short video walkthrough guide on the toolkit website.

The example we'll use here is a fictional study exploring whether changing the wording of a letter can increase the number of parents applying for the 2 year old free childcare offer. The steps would be similar when you run your own study.

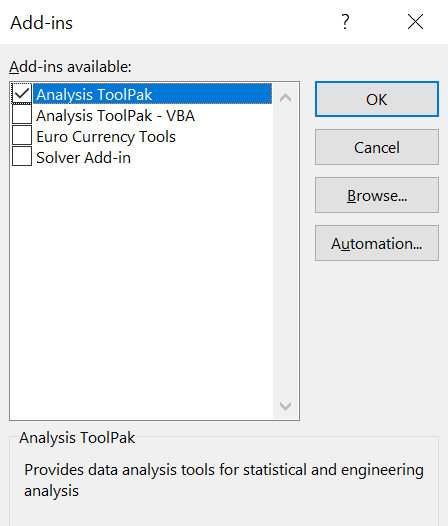
### **Step one:** Installing the analysis add-on

To conduct the analysis we need to install an add-on (‘Analysis ToolPak’). Click the ‘File’ tab at the top of your screen and select ‘Options’ (you may need to select ‘More’ at the bottom to reach Options). A panel like this will appear.



Click the ‘Add-ins’ tab on the left. Go down to the ‘Manage Excel Add-ins’ menu at the bottom and click ‘Go’.

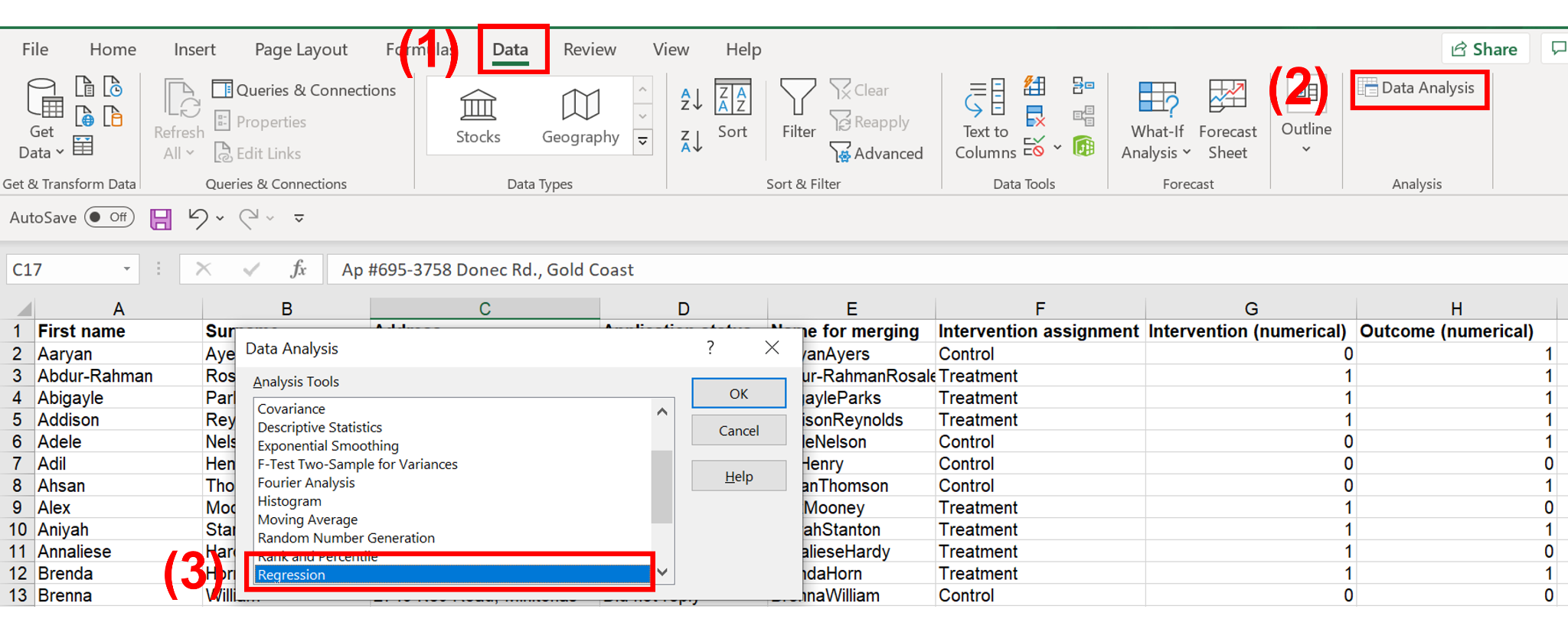
Download the ‘Analysis ToolPak’ by checking the Analysis ToolPak check box, and clicking OK.



### **Step two:** Analysing your data using regression

Now we can analyse our data. Select your whole dataset using the column headers. Click the ‘Data’ tab at the top of the screen.

Now scroll to the right to find the ‘Data analysis’ button and select ‘Regression’ from the list.

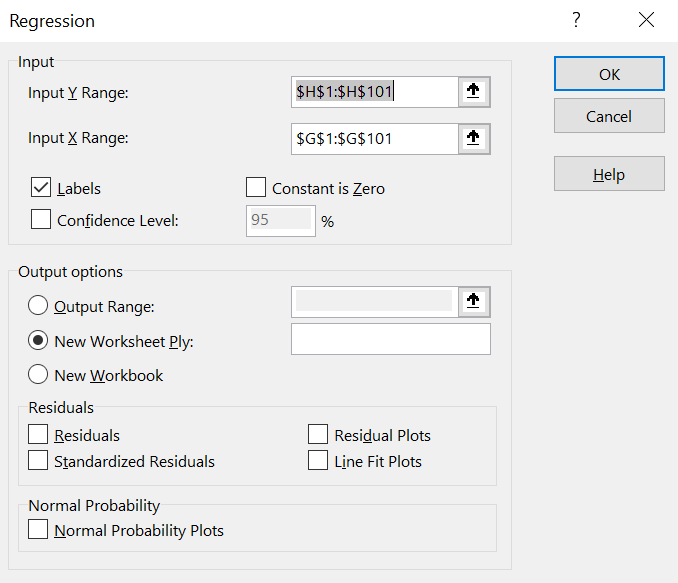


Type the cell references for your outcome data into the ‘Input Y Range’ bar. For the example in the Excel guide, the reference is H1:H101.

Now do the same for your randomisation data in the ‘Input X range’ bar. For us, this is G1:G101.

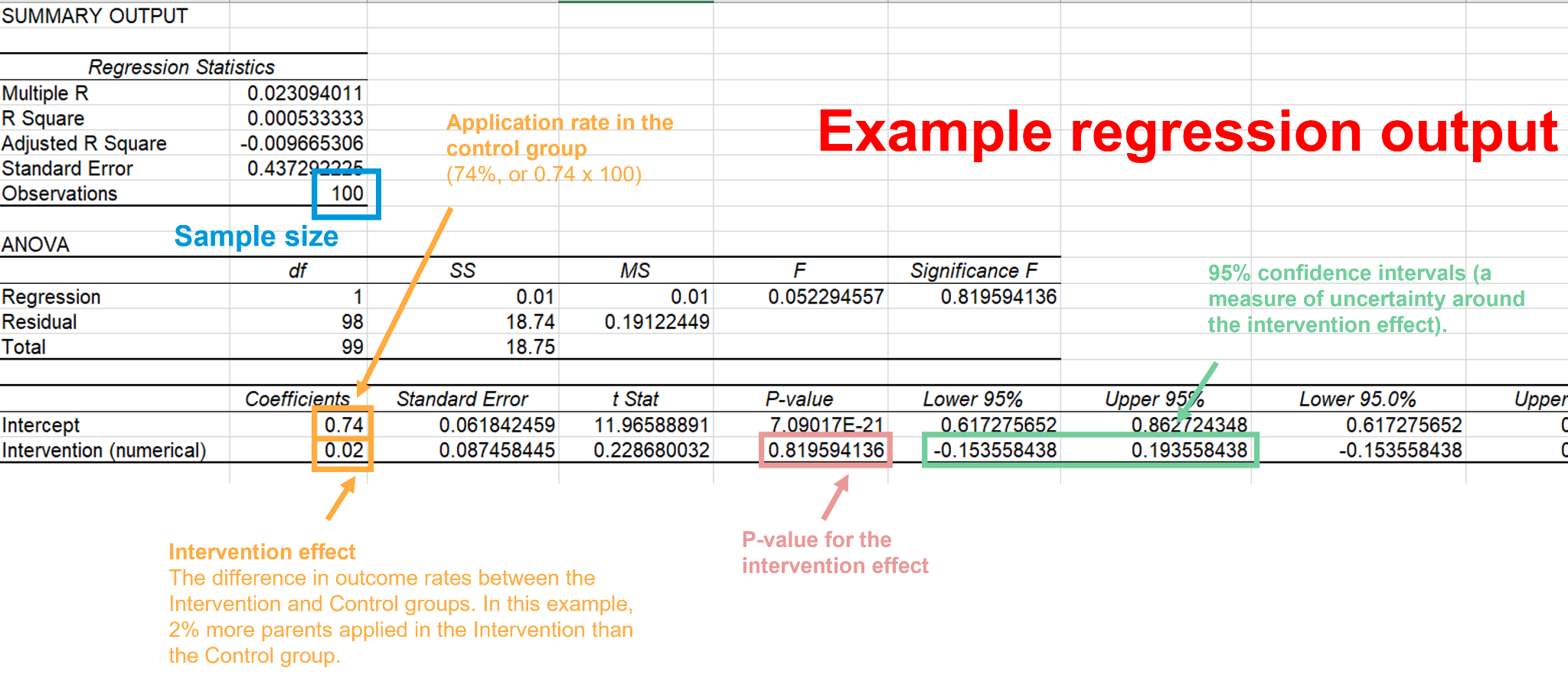
Both of these ranges need to be formatted as 1’s and 0’s, not words, for Excel to use them for analysis (see the *data cleaning and merging guide* for instructions on how to change from one to the other).

**Note:** Make sure the “Labels” box is ticked to include column headings.



Once you click “OK”, Excel will create a new worksheet with the regression output. This can look pretty confusing, but you only need to pay attention to a few parts.

In this regression output, the main things to look at are highlighted and colour coded in the screenshot below:



Going into a bit more detail about each of them:

* The “Intercept” coefficient (the number in the upper **orange** box in the “Coefficients” section). This is the application rate in the Control group. We multiply this coefficient of 0.74 by 100 to get the percentage value of 74%.
* The “Intervention (numerical)” coefficient (the number in the lower **orange** box), which is the effect of your intervention itself. This is the difference between the Treatment and Intervention group’s application rates. In this case, a value of 0.02 means that the Intervention group has an application rate that is 2% higher than the Control group (0.02 times 100 to get the percentage). We can also use this number to get the application rate in the Intervention group, which is 76% (74% + 2%).

* The p-value (the number in the **pink** box). This tells us if your policy has had a statistically significant effect, which is the case if the p-value is less than 0.05. Here the p-value is 0.8, which means the effect is not statistically significant. The main toolkit contains further information on how to interpret p-values.
* The numbers in the **green** box are the “95% confidence intervals” around the effect of your intervention. This is a measure of the level of uncertainty, with *smaller confidence intervals meaning less uncertainty (better).* Specifically, if the study was repeated many times, 95% of the time the value of the effect would be somewhere within this interval. In this case, it means that 95% of the time the effect would be somewhere between -0.15 and 0.19 (i.e. -15% and 19%, a wide range, indicating high uncertainty).
* The number in the **blue** box is the number of people in our sample.

### **Step three**: Generating your results graph

To create a graph we will insert values from the regression output.

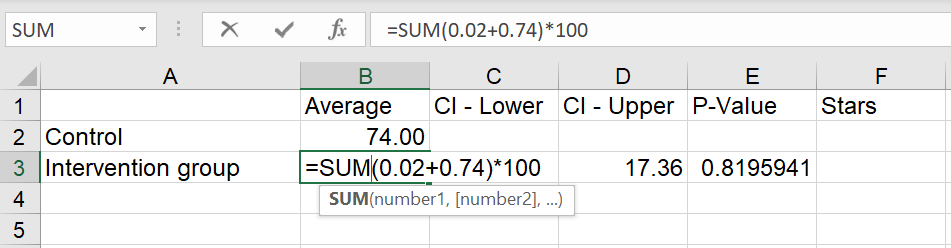
First, go to the ‘3. Graphing template’ sheet. You’ll see an example of what a full results graph looks like. There are 4 main elements to the graph:

1. The blue and black results bars
2. Confidence intervals
3. Statistical significance
4. Sample size and formatting

**1 (Blue and black results bars):** In column B you can see the application rates for each of the two groups (control and intervention). We can find this information in the regression output’s “Coefficients” column (the cells in the **orange** box in the screenshot above):

* The “Intercept” row is the application rate in the Control group. We multiply the coefficient of 0.74 by 100 to report the percentage value of 74% for the Control group on the graph.
* The “Intervention (numerical)” row is the difference between the Control and the Intervention group. In this case, a value of 0.02 means that the Intervention group has an application rate that is 2% higher than the Control group (0.02 times 100 to get the percentage). So instead of directly reporting this coefficient value on the graph, we add the 2% to the Control group’s average of 74% to get an application rate of 76% in the Intervention group. You can do this using the Excel command:

=SUM(0.02+0.74)\*100



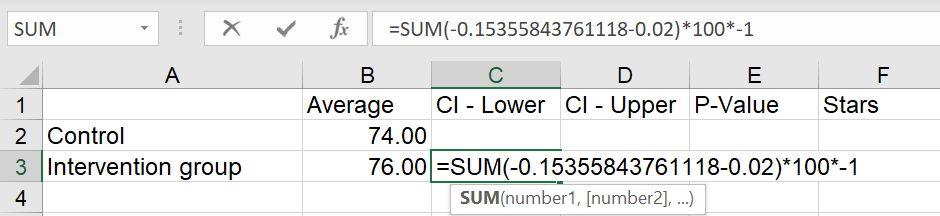
When you do this with your own data, replace the values with whatever values come out of your regression output.

Once both cells are filled out, the graphing template will automatically adjust the height of the graphs’ two bars.

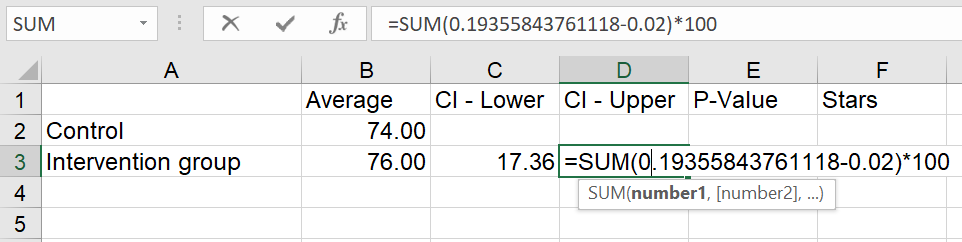
**2 (Confidence intervals):** The graphing template’s Columns C and D titled “CI - Lower” and “CI - Upper” refer to the lower and upper 95% confidence intervals of the intervention effect, which is the orange line visible around the blue bar. This is a measure of the level of uncertainty around an average value, with *smaller* confidence intervals meaning *less uncertainty* (i.e. better).

To calculate your confidence interval:

1. Take the value from the ‘Lower 95%’ column and the ‘Intervention (numerical)’ row of your regression output (in the **green** box in the summary output screenshot above). Now type ‘=SUM(‘, then copy in the value. Deduct the value of the ‘Intervention (numerical)’ coefficient from it, and multiply the whole thing by -100.



1. Do the same for the ‘Upper 95%’ value from the ‘Intervention (numerical)’ row (also in the **green** box above). This time, multiply by 100 instead of -100.



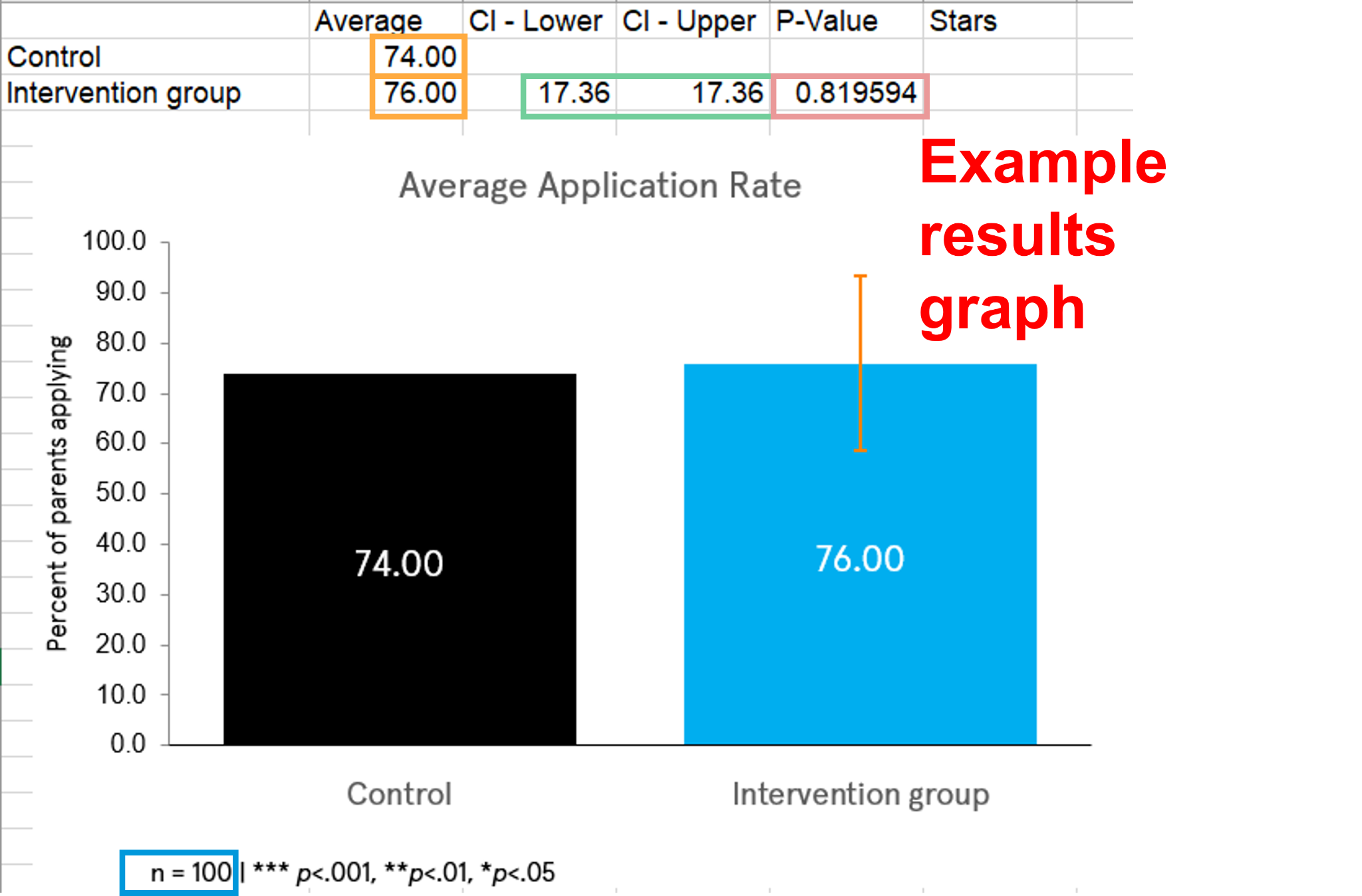
In many analyses, the intervals are symmetrical, so the two values are identical (17.36), but this is not always the case. Once both “CI” cells are filled out, the graphing template will automatically adjust the length of the confidence interval lines in the graph.

**3 (Statistical significance):** Finally, add your p value. You can find this in the ‘P-value’ column and the ‘Intervention (numerical)’ row (in the **pink** box in the summary output screenshot above).

If your result is statistically significant, a star will automatically appear on the graph. The number of asterisks that appear on the graph will automatically vary depending on the magnitude of the p-value. This information is also explained in the footnote of the graph.

**4: (Sample size and formatting):** Once your graph is ready, you might want to work on the final touches. This includes changing the graph title, adjusting the range of the vertical axis, and updating the sample size, reported as “n” on the footnote (we can find this value on the output table’s “Observations” cell in the **blue** box in the screenshot above).

Once you’ve done all this, you’ll end up with a results graph that looks something like this. The coloured boxes here correspond to the same coloured boxes on the regression output ( the confidence intervals in the **green** box below will be slightly different to those in the regression output, because we transformed them in step 3).



That’s it! Your results graph is ready and you’ve finished analysing your data. Now you can do this for your own dataset.