Instructions for using the Excel Bootstrap Tool: One data set

14 July 2021

Entering data

Your data goes into Columns B (Quantities) and C (Weights) starting in Row 3. A Quantity is the data value from a single recording. A Weight is the length of time of each recording. For example, if you want to compute the mean open time (MOT) across 10 recordings, fill B3 with the MOT from Recording 1 and C3 with the length of Recording 1. Fill in B4,…,B12 with the MOTs from the other recordings and C4,…C12 with the respective recording lengths. As you fill Column B, Column A will autopopulate with a running index of the record number. This is just to better keep track of how many recordings you have entered data for.

Note on weights. You can choose whatever weighting you want for the data (it does not have to be record length). If you want all data weighted equally, put 1 in each Weight cell. The weights can be in any units, as long as they are the same for all entries. The weights do not have add up to 1; the calculation takes care of that automatically.

There can be a lag of a few seconds between updating a cell and the recalculation of the bootstrapping. This can be annoying when entering data, so the easiest ways to enter data without this lag is either:

1. cut and paste the data from another application or Excel spreadsheet. You should see the bootstrap data update. If the bootstrap data does not compute automatically, then *either* select Formulas > Calculation Options > Automatic *or* click on Calculate Now in the Formulas ribbon.
2. turn off the automatic recalculation of all formulas (Formulas > Calculation Options > Manual). To update the bootstrap data after entering the data, click on Calculate Now in the Formulas ribbon.

Important: If you want to get rid of old data in Columns B and C, do so by selecting the existing data (click on a corner cell and drag the mouse until the rectangle contains the data) and then pressing Backspace or Delete (or right-clicking and choosing Clear Contents). **DO NOT** use the Excel Delete function located in the Home ribbon to get rid of old data. Selecting and pressing the Delete *key* just clears the numbers from the cells (GOOD), while using the Excel Delete *function* in the Home ribbon deletes underlying metadata that the worksheet needs (BAD). If other cells physically move after deletion or if you are asked which way cells should shift, you are using the Delete *function* that you shouldn't be using. If all the calculated cells become #REF!, then you probably deleted cell B3 in a bad way. Use Control-Z to undo the problem. (Followed by Calculate Now, if you're using that.)

Bootstrap output

All the output data you need is in Sheet 1 of the spreadsheet. The columns in Sheet 1 contain the following:

Column L. This is a list of all the bootstrap means that were created. The length of this equal to the number in Column E. This is of interest if you want to plot a histogram of the bootstrap mean, but this is done for you.

Column E. This shows the number of bootstrap calculations that were done. The program computes 50,000 random samples (in Sheet 2) and then groups those into batches that have the length of your data in Column B. If this number of data points is n, then the number of bootstrap means generated is $50{,}000/n$.

Column F. This is the weighted average of your data (Column B with weights in Column C). It is computed as the mean of the bootstrap means in Column L.

Column G. The data's standard error of the mean (equal to the standard deviation (SD) of the bootstrap means).

Columns H and I. These two numbers make up the 95% confidence interval, the suggested CI for bootstrap. Column G is the lower bound (the 2.5% percentile, aka 0.025 quantile, of the bootstrap means in Column L). Column H is the upper bound (the 97.5% percentile, aka 0.975 quantile).

Columns J and K. These are the 75% confidence interval (the 12.5% and 87.5% percentiles of Column L).

The histogram is of the list of bootstrap means in Column L as a probability distribution (i.e., counts/#bins/bin width). It uses 50 bins. The frequencies plotted in the histogram are computed in Sheet3. The red line is a Gaussian with the same mean and SD as the bootstrap means to visually assess how close the histogram is to a Gaussian.

We will keep an updated version of the program in the Harvard Dataverse repository at https://doi.org/10.7910/DVN/RTFGBG. Comments/questions: dirk_gillespie@rush.edu