Chapter 2

# Try It Yourself: Impute Missing Values from Internal Data

## Excel Instructions

1. Open the vote.csv data set in Excel.
2. Verify that gpa is in column D.
3. Calculate mean and median before imputation:
   1. Type the formula:

=AVERAGE(D:D) in Cell J1

* 1. Type “Mean before imputation” in Cell I1.
  2. Type the formula:

=MEDIAN(D:D) in Cell J2

* 1. Type “Median before imputation” in Cell I2.

1. Perform mean imputation:
   1. Highlight the E column to the right of gpa and Right-Click → Insert.
   2. Title this new column “imputed\_gpa” in cell E1.
   3. Use an IF statement combined with ISBLANK. Assuming the original gpa is in column D, in the first row of the new column (cell E2), enter: =IF(ISBLANK(D2), K$1, D2).
   4. Apply this formula to all rows of this column.
2. Calculate mean and median after imputation:
   1. Type the formula:

=AVERAGE(E:E) in Cell K4

* 1. Type “Mean before imputation” in Cell J4.
  2. Type the formula:

=MEDIAN(E:E) in Cell K5

* 1. Type “Median before imputation” in Cell J5.