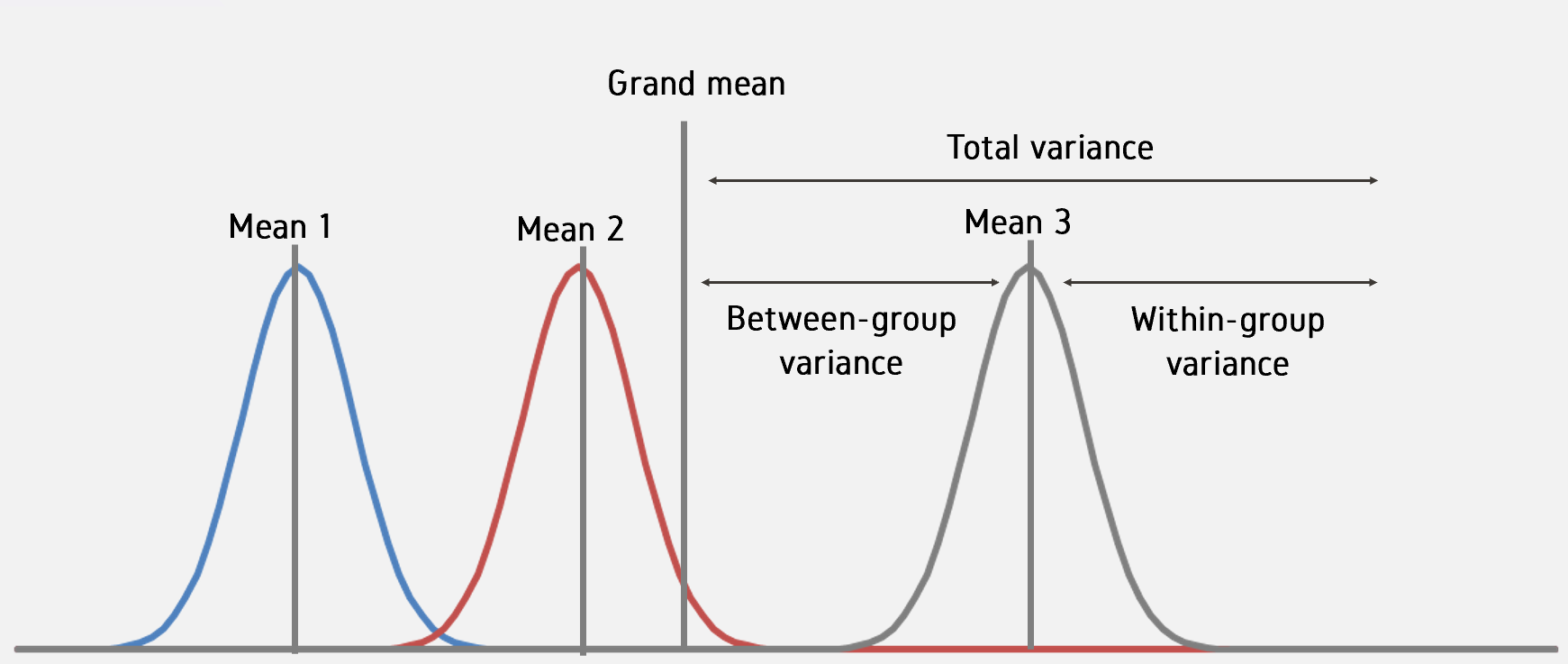
**Analysis of variance (ANOVA)**

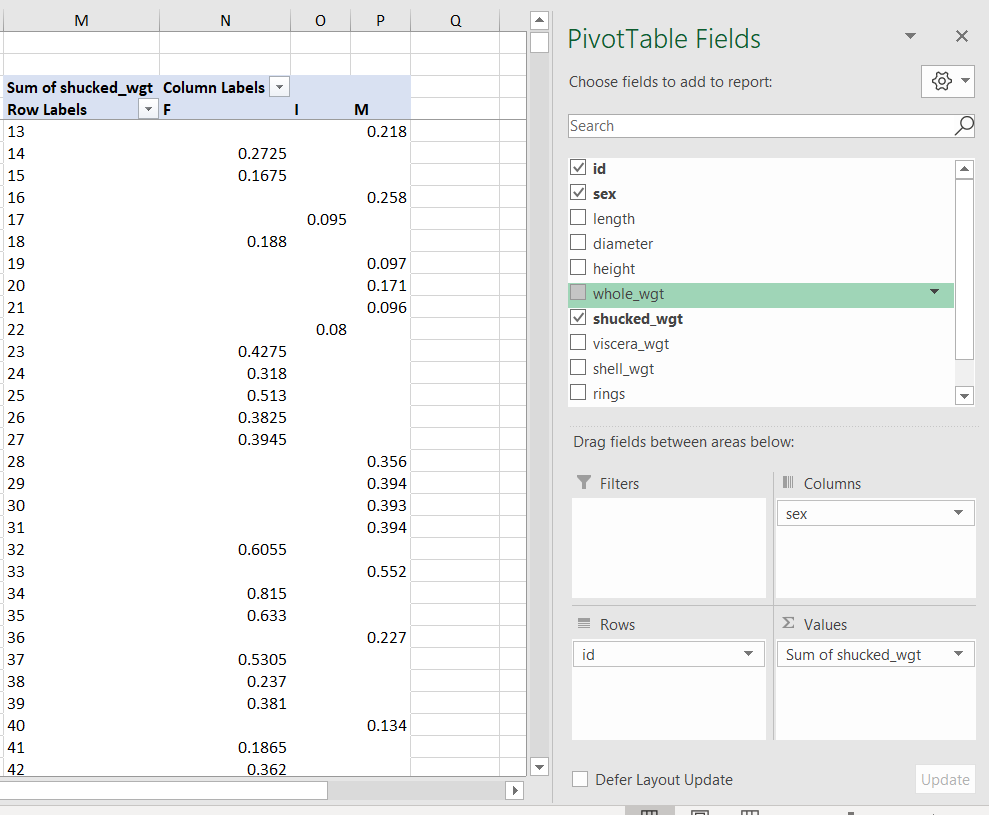
File: datasets/abalone.xlsx  
Solution: solutions/anova-solution.xlsx

The analysis of variance, or ANOVA, is used to compare the difference in means across more than two groups. The goal is to measure the ratio of *between-group variance* and *within-group variance*. If there’s more within-group variance, that means one group has a significantly different mean.

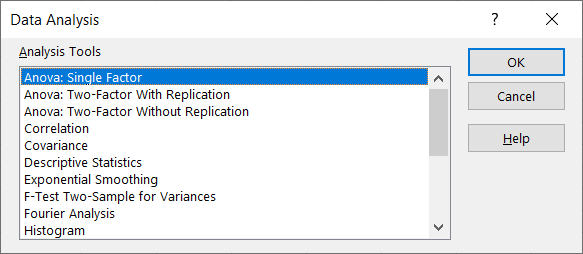


Let’s check for a significant difference in shucked weights across male, female and infant snails.

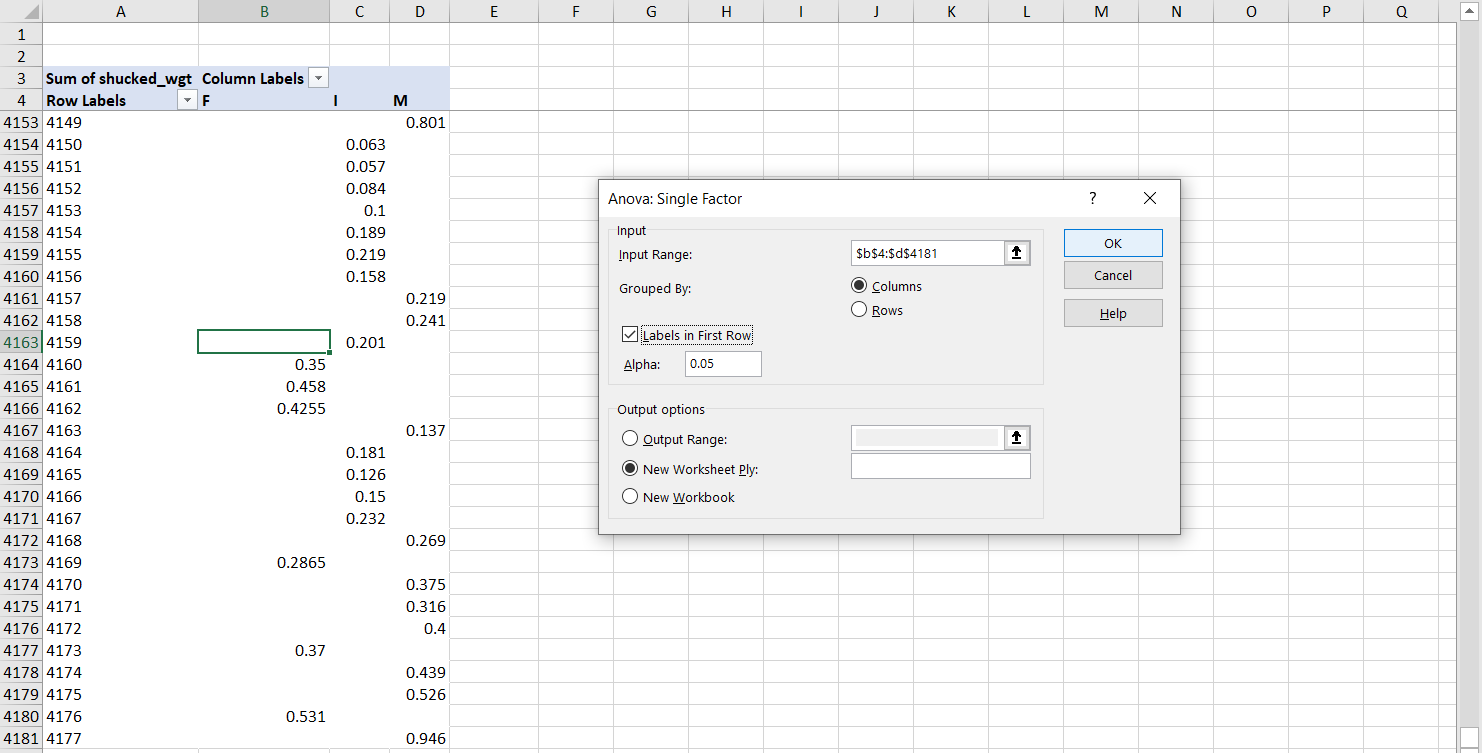
1. Insert a PivotTable. Put id in the Rows section, sex in the Columns section and Sum of shucked\_wgt in the Values section.
   1. Turn off the totals by clicking inside the PivotTable and selecting Design > Grand Totals > Off for Rows and Columns.



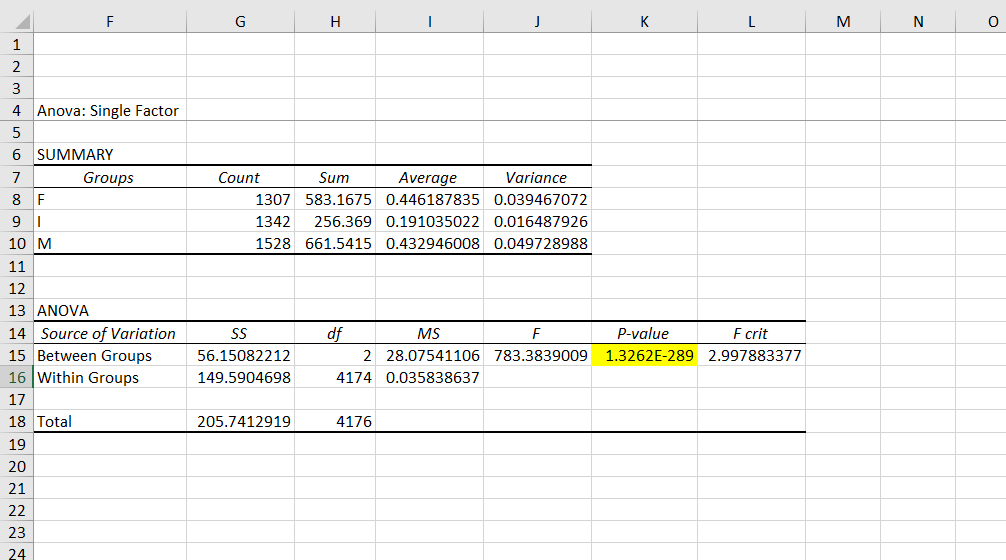
1. In the Analysis ToolPak, select Anova: Single Factor.



* 1. The input range is the three columns for each category: F, I and M.



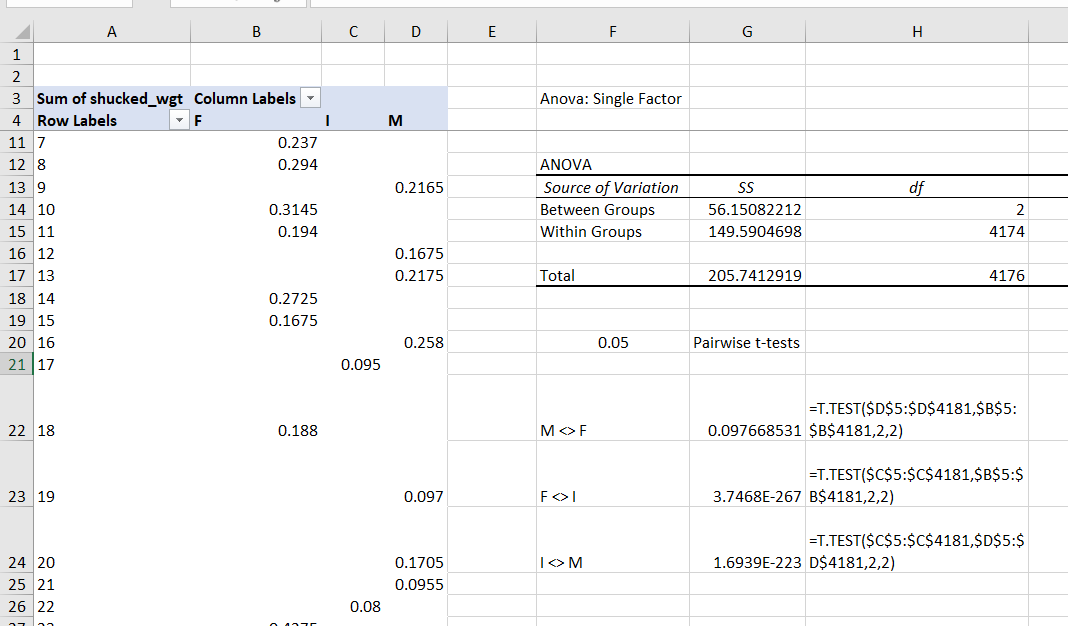
1. The results of the ANOVA are available in the second box of outputs. The p-value for between-groups variation tells us if there is a significant difference across group means.



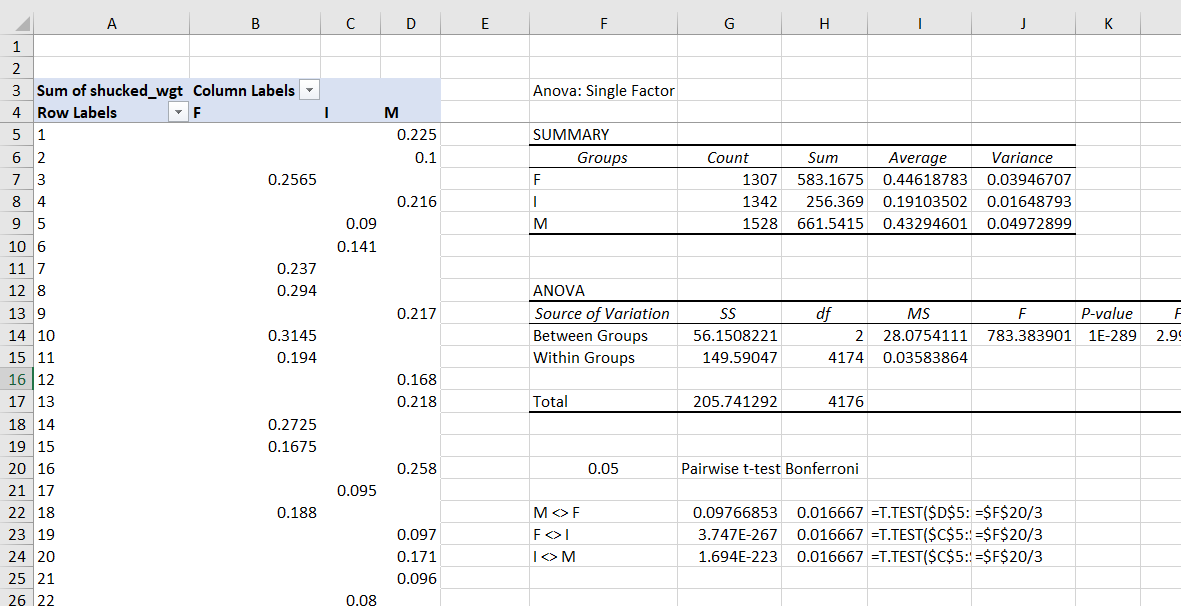
**ANOVA post-hoc tests: pairwise comparisons with Bonferroni correction**

The ANOVA in itself does *not* tell us *which* groups in particular are significantly higher/lower than the others. To do that, we will run *post-hoc* tests while adjusting for *experimentwise error*… that is, the more tests you run, the more likely you are to find a significant p-value just out of chance.

1. Conduct a *pairwise t-test* for to compare each pair of categories using the T.TEST() function.
   1. This will take four arguments:
      1. The range containing the first category to compare
      2. The range containing the second category to compare
      3. Whether this is a one- or two-tail test. We are using two-tail tests, so the argument is 2.
      4. The type of t-test. Since these are independent samples, this is not a paired t-test. We will assume equal variances as that is an assumption of the ANOVA. So the argument here is also 2.
      5. The result of T.TEST() is the test’s p-value. We will compare it against the adjusted alpha next.



1. We will now compare these p-values to a Bonferroni-adjusted alpha. This number will be our original alpha (.05) divided by the number of groups we are comparing (3). This makes our new alpha .0017.



* 1. Based on these results, there is no significant difference in weights between male and female snails, but there is a significant difference between female and infant snails, and a significant difference between male and infant snails.