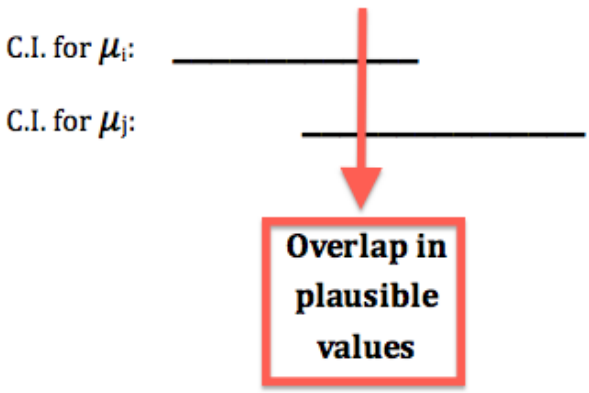
ANOVA: Summary

Learning Objective: In a given context, carry out the inferential method for comparing groups and draw the appropriate conclusions.

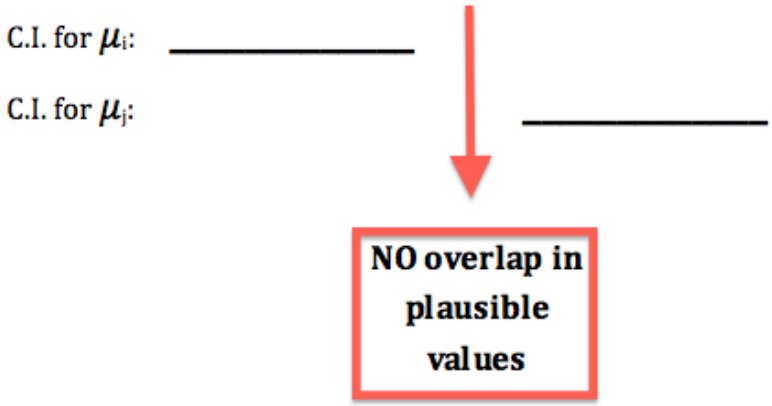
Final Comment

However, the ANOVA F-test does not provide any insight into why H_0 was rejected; it does not tell us in what way $\mu_1, \mu_2, \mu_3 \ldots, \mu_k$ are not all equal. We would like to know which pairs of 's are not equal. As an exploratory (or visual) aid to get that insight, we may take a look at the confidence intervals for group population means$\mu_1, \mu_2, \mu_3 \ldots, \mu_k$ that appears in the output. More specifically, we should look at the position of the confidence intervals and overlap/no overlap between them.

* If the confidence interval for, say,μ_i overlaps with the confidence interval for μ_j , then μ_i and μ_j share some plausible values, which means that based on the data we have no evidence that these two 's are different.



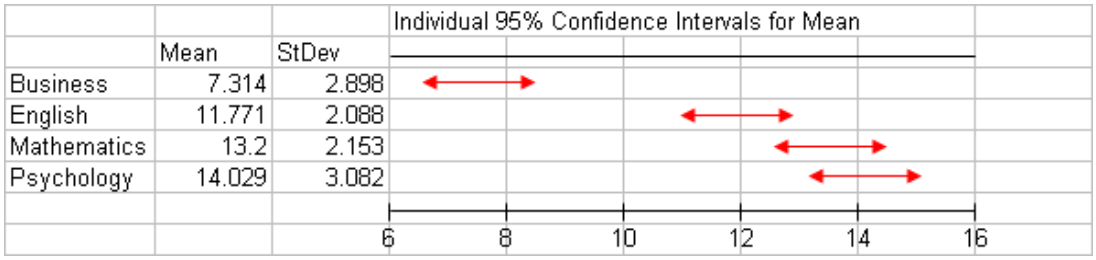
* If the confidence interval for μ_i does not overlap with the confidence interval for μ_j , then μ_i and μ_j do not share plausible values, which means that the data suggest that these two 's are different.



Furthermore, if like in the figure above the confidence interval (set of plausible values) for μ_i lies entirely below the confidence interval (set of plausible values) for μ_j, then the data suggest that μ_i is smaller than μ_j.

Example

Consider our first example on the level of academic frustration.



Based on the small p-value, we rejected H_0 and concluded that not all four frustration level means are equal, or in other words that frustration level is related to the student's major. To get more insight into that relationship, we can look at the confidence intervals above (marked in red). The top confidence interval is the set of plausible values for μ_1, the mean frustration level of business students. The confidence interval below it is the set of plausible values for μ_2, the mean frustration level of English students, etc.

What we see is that the business confidence interval is way below the other three (it doesn't overlap with any of them). The math confidence interval overlaps with both the English and the psychology confidence intervals; however, there is no overlap between the English and psychology confidence intervals.

This gives us the impression that the mean frustration level of business students is lower than the mean in the other three majors. Within the other three majors, we get the impression that the mean frustration of math students may not differ much from the mean of both English and psychology students, however the mean frustration of English students may be lower than the mean of psychology students.

Note that this is only an exploratory/visual way of getting an impression of why H_o was rejected, not a formal one. There is a formal way of doing it that is called "multiple comparisons," which is beyond the scope of this course. An extension to this course will include this topic in the future.

Let's Summarize

- The ANOVA F-test is used for comparing more than two population means when the samples (drawn from each of the populations we are comparing) are independent. We encounter this situation when we want to examine the relationship between a quantitative response variable and a categorical explanatory variable that has more than two values.
- The hypotheses that are being tested in the ANOVA F-test are: $H_0 : \mu_1 = \mu_2 = \ldots = \mu_k$
  $H_a : not\ \ all\ \ the\ \ \mu's\ \ are\ \ equal$
- The idea behind the ANOVA F-test is to check whether the variation among the sample means is due to true differences among the μ's or merely due to sampling variability by looking at:
  $$\frac{Variation\ \ among\ \ the\ \ sample\ \ means}{Variation\ \ within\ \ the\ \ groups}$$
- Once we verify that we can safely proceed with the ANOVA F-test, we use software to carry it out.
- If the ANOVA F-test has rejected the null hypothesis we can look at the confidence intervals for the population means that are in the output to get a visual insight into why H_o was rejected (i.e., which of the means differ).
