

⚠ Lagunita is retiring and will shut down at 12 noon Pacific Time on March 31, 2020. A few courses may be open for self-enrollment for a limited time. We will continue to offer courses on other online learning platforms; visit <http://online.stanford.edu>.

Course > Inference: Relationships C→Q > ANOVA > ANOVA: The Idea behind the F-Test

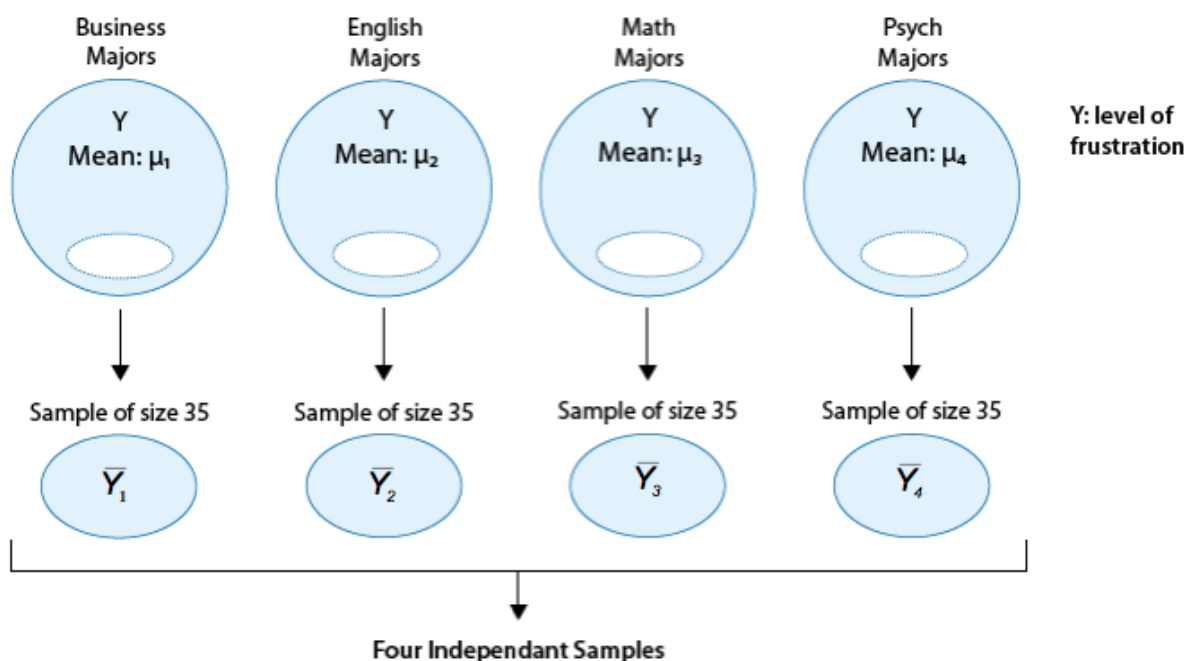
🔖 Bookmark this page

ANOVA: The Idea behind the F-Test

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

The idea behind the ANOVA F-Test

Let's think about how we would go about testing whether the population means $\mu_1, \mu_2, \mu_3, \mu_4$ are equal. It seems as if the best we could do is to calculate their point estimates—the sample mean in each of our 4 samples (denote them by $\bar{y}_1, \bar{y}_2, \bar{y}_3, \bar{y}_4$),



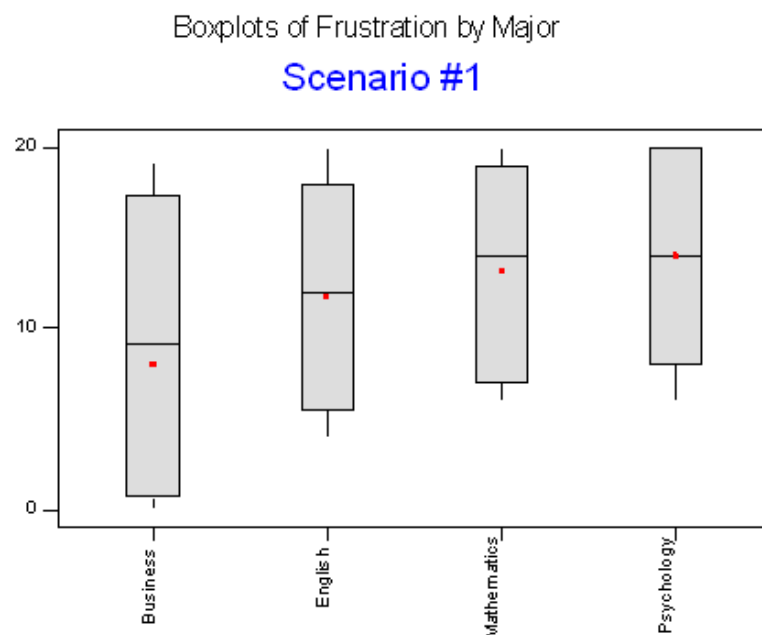
and see how far apart these sample means are, or in other words, measure the variation between the sample means. If we find that the four sample means are not all close together, we'll say that we have evidence against H_0 , and otherwise, if they are close together, we'll say that we do not have evidence

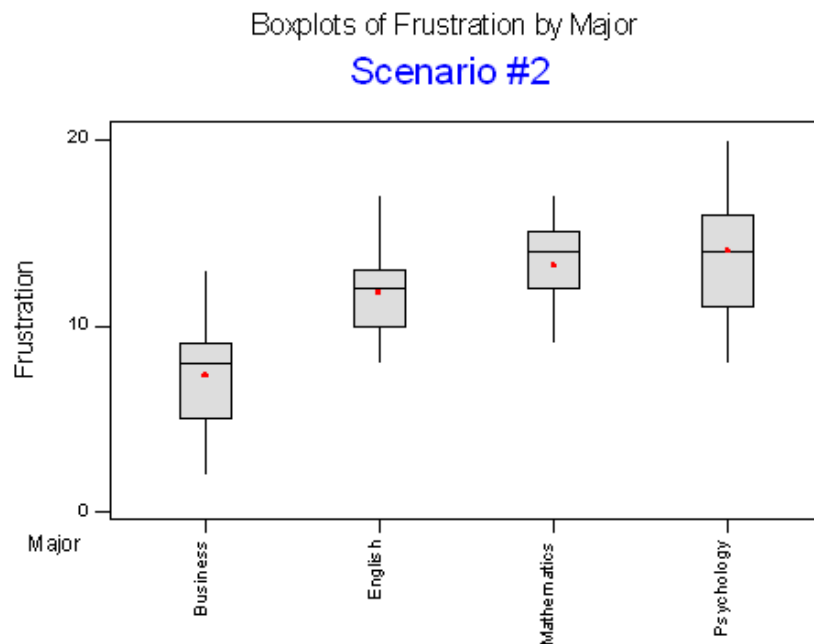
against H_0 . This seems quite simple, but is this enough? Let's see.

It turns out that:

- * The sample mean frustration score of the 35 business majors is: $\bar{y}_1 = 7.3$
- * The sample mean frustration score of the 35 English majors is: $\bar{y}_2 = 11.8$
- * The sample mean frustration score of the 35 math majors is: $\bar{y}_3 = 13.2$
- * The sample mean frustration score of the 35 psychology majors is: $\bar{y}_4 = 14.0$

Below we present two possible scenarios for our example. In both cases, we construct side-by-side boxplots for four groups of frustration levels that have the same variation among their means. Thus, Scenario #1 and Scenario #2 both show data for four groups with the sample means 7.3, 11.8, 13.2, and 14.0 (indicated with red marks).





Learn By Doing

1/1 point (graded)

Look carefully at the graphs of both scenarios. For which of the two scenarios would you be willing to believe that samples have been taken from four groups which have the same population means?

☒ Scenario 1 ✓

☐ Scenario 2

Answer

Correct: Your intuition is right. Read on to understand what's going on here.

Submit

The important difference between the two scenarios is that the first represents data with a large amount of variation within each of the four groups; the second represents data with a small amount of variation within each of the four groups.

Scenario 1, because of the large amount of spread within the groups, shows boxplots with plenty of overlap. One could imagine the data arising from 4 random samples taken from 4 populations, all having the same mean of about 11 or 12. The first group of values may have been a bit on the low side, and the other three a bit on the high side, but such differences could conceivably have come about by chance. This would be the case if the null hypothesis, claiming equal population means, were true. Scenario 2, because of the small amount of spread within the groups, shows boxplots with very little

overlap. It would be very hard to believe that we are sampling from four groups that have equal population means. This would be the case if the null hypothesis, claiming equal population means, were false.

Thus, in the language of hypothesis tests, we would say that if the data were configured as they are in scenario 1, we would not reject the null hypothesis that population mean frustration levels were equal for the four majors. If the data were configured as they are in scenario 2, we would reject the null hypothesis, and we would conclude that mean frustration levels differ, depending on major.

Let's summarize what we learned from this. The question we need to answer is: Are the differences among the sample means (\bar{Y} 's) due to true differences among the μ 's (alternative hypothesis), or merely due to sampling variability (null hypothesis)?

In order to answer this question using our data, we obviously need to look at the variation among the sample means, but this alone is not enough. We need to look at the variation among the sample means relative to the variation within the groups. In other words, we need to look at the quantity:

VARIATION AMONG SAMPLE MEANS

VARIATION WITHIN GROUPS

which measures to what extent the difference among the sampled groups' means dominates over the usual variation within sampled groups (which reflects differences in individuals that are typical in random samples).



When the variation within groups is large (like in scenario 1), the variation (differences) among the sample means could become negligible and the data provide very little evidence against H_0 . When the variation within groups is small (like in scenario 2), the variation among the sample means dominates over it, and the data have stronger evidence against H_0 .

Looking at this ratio of variations is the idea behind the comparing more than two means; hence the name analysis of variance (ANOVA).

Now that we understand the idea behind the ANOVA F-test, let's move on to step 2. We'll start by talking about the test statistic, since it will be a natural continuation of what we've just discussed, and then move on to talk about the conditions under which the ANOVA F-test can be used. In practice, however, the conditions need to be checked first, as we did before.

Open Learning Initiative 



 Unless otherwise noted this work is licensed under a Creative Commons Attribution-NonCommercial-ShareAlike 4.0 International License .

