Oneway Analysis of Variance (ANOVA)

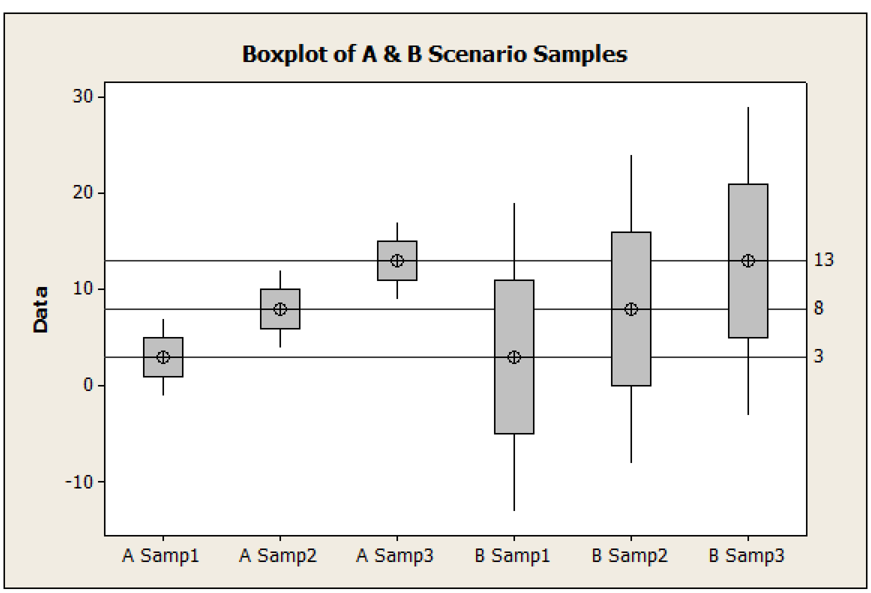
Overview: Oneway ANalysis Of VAriance (ANOVA) is a method for comparing the means of I different populations or groups. (For example, we might be interested in comparing the compression strength of a certain type of concrete cured I different ways.) The name is derived from the fact that we test the **ANOVA hypotheses**,

H_0: $\mu_1 = \mu_2 = \cdots = \mu_I$ vs.
H_1: *at least two means differ*

by comparing two variance estimates, **MSE** and **MSTr**, via the **F-ratio F = MSTr/MSE**.

ANOVA vs. Multiple Two-Sample Tests: Suppose we want to test H_0: $\mu_1 = \mu_2 = \mu_3$. Why not simply test H_{0_1}: $\mu_1 = \mu_2$, H_{0_2}: $\mu_2 = \mu_3$, and H_{0_3}: $\mu 1 = \mu_3$ using three two-sample tests? In other words, why not test the ANOVA null hypothesis by testing the equality of every possible pair of means? The reason we don't is that it would result in inflated type I error probabilities. For example, if you conduct three independent tests at $\alpha = 0.05$, the probability of committing at least one type I error is about 0.14. If you conduct 10 tests at $\alpha = 0.05$ the probability of at least one type I error is about 0.40. These probabilities are unacceptably high. To circumvent this problem we need to test the ANOVA hypotheses using a **single test**. The F-test provides a way to do this.

The Logic of the F-test: As mentioned above, the F-ratio test statistic is a ratio of two measures of variance: F = MSTr/MSE. To illustrate the logic behind this ratio, we consider two scenarios (A, B) comparing three means. Boxplots for the two scenarios are presented below with scenario A on the left and scenario B on the right. Note that the mean of each sample is represented by a circle. Which scenario provides the greater evidence against H_0: $\mu_1 = \mu_2 = \mu_3$?



Clearly scenario A provides greater evidence against H_0 than scenario B. But note that the sample means in the two scenarios are the same, that is, the between-sample-mean variation in both scenarios is the same. *But*, the within-sample variation is much greater for scenario B. The IQR's for the scenario B samples are four times greater than those for the scenario A samples. Thus we see that the ratio of between-sample variation to within-sample variation is much greater for A than for B, sixteen times greater since variance is the square of the standard deviation, our usual measure of spread. But, properly scaled, this is exactly the ratio measured by the F-ratio: the F-ratio for A's data is $F = 25$ vs $F = 1.5625$ for B's data, i.e., it is 16 times greater.

Interpreting the F-ratio. From the preceding discussion we see that the larger the F-ratio the greater the evidence against the null hypothesis H_0: $\mu_1 = \mu_2 = \cdots = \mu_I$ and in favor of the alternative hypothesis H_1: *at least two means differ*. Since large values of F (F >> 1) provide evidence against H_0 the p-value for testing H_0 vs H_1 is given by $P(F \geq F_{\text{actual}})$, computed assuming H_0 true, where F_{actual} is the F-ratio for the actual sample data. In order to compute the p-value we need the null (H_0 true) distribution of F, which is the F-distribution. We will discuss this plus the formula for computing the F-ratio later. We finish by briefly discussing the Minitab `Stat -> ANOVA -> One-Way` output for the two scenarios. Below we provide Scenario A's output followed by scenario B's:

```
One-way ANOVA: A Samp1, A Samp2, A Samp3

Source  DF      SS      MS      F      P
Factor   2  350.00  175.00  25.00  0.000
Error   18  126.00    7.00
Total   20  476.00

One-way ANOVA: B Samp1, B Samp2, B Samp3

Source  DF    SS   MS     F      P
Factor   2   350  175  1.56  0.237
Error   18  2016  112
Total   20  2366
```

Note that the calculations for the F ratio for each scenario are presented in what is called an ANOVA table. At the far right of the first row of each table we see that the F-ratio is provided along with the corresponding p-value. Since the F-ratio for scenario A is much larger than that for scenario B, the p-value for scenario A is much smaller than that for scenario B, approximately 0 vs 0.237. So we reject the null hypothesis of all means equal for scenario A's data but not for scenario B's.