Oneway ANOVA

This figure depicts the data for an experiment in which 150 4th graders were randomly assigned to three groups. One group received a modest visual sightsinging training treatment (i.e., weak), one group received an intense visual sightsinging training treatment (i.e., strong), and one group received typical verbal instruction (i.e., control). The outcome measure was sightsinging achievement. Compare the list of notes to the figure side-by-side:

Observe the following regarding the variation in the data

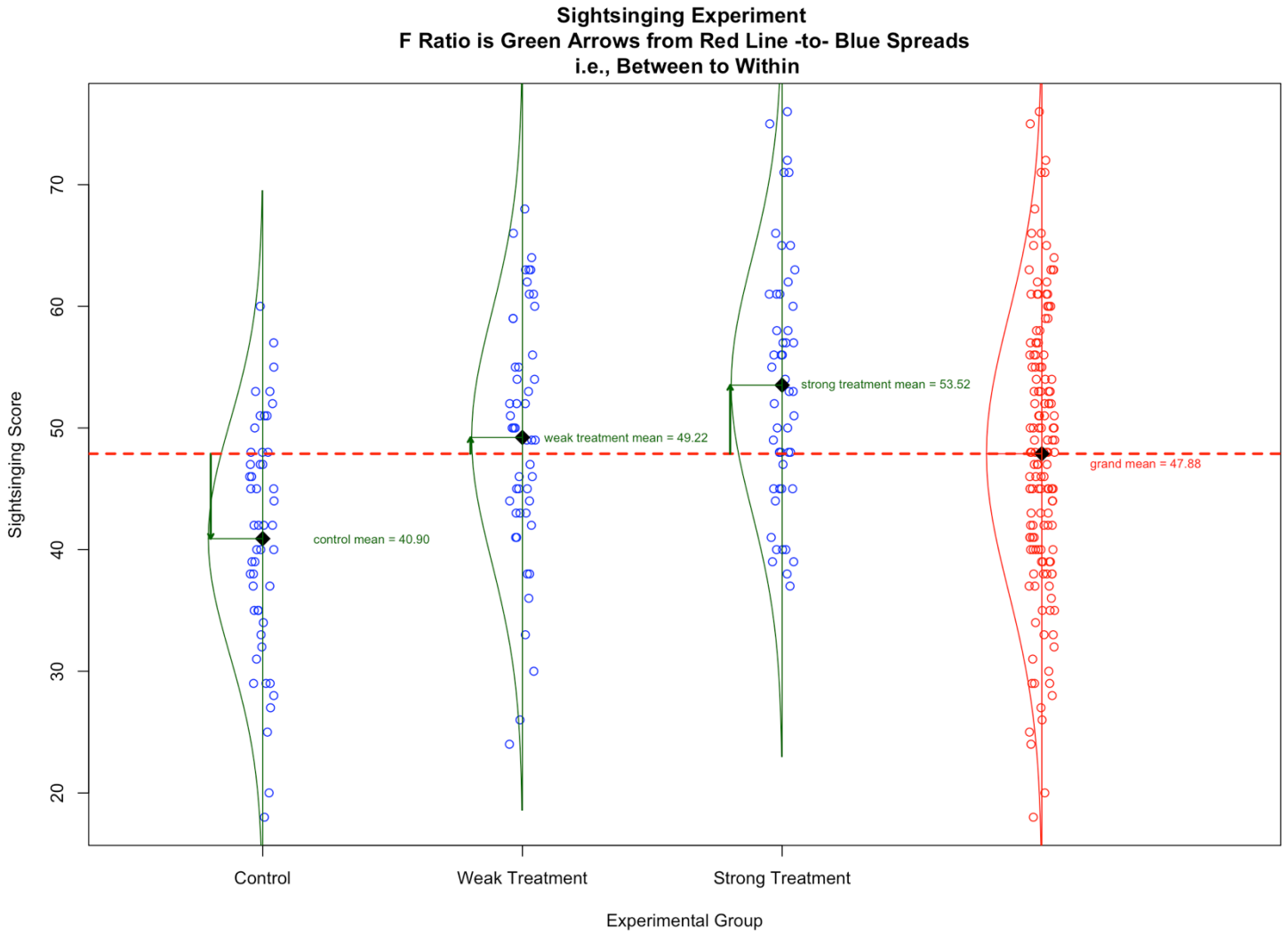
- The cluster of red dots at the far right is the variation among the 150 participants on sightsinging achievement overall, WITHOUT any grouping. This is the visualization of the TOTAL VARIATION – in other words this is the variation represented by $\mathbf{SS_{total}}$
- Oneway ANOVA "partitions" or "decomposes" the TOTAL VARIATION into WITHIN GROUPS VARIATION and BETWEEN GROUPS VARIATION.
- The clusters of blue dots represent the WITHIN GROUPS VARIATION (e.g., error/residual variation) – in other words, if you calculated the weighted variation within each group according to their respective means and totaled them, this is the variation represented by $\mathbf{SS_{within}}$
- The variation of the group means from the **"grand mean"** (e.g., **the green arrows from the dotted red line**, ***27.19***) represents the BETWEEN GROUPS VARIATION (e.g., model variation) – in other words, if you calculated the average weighted differences of the group means and the grand mean, this is the variation represented by $\mathbf{SS_{between}}$
- *Assumptions:* The fact that any given blue data point only appears in one group represents the model assumption of independence. The normal curves super-imposed on each distribution of participants by GROUP represents the model assumption of normality, the equivalency in the dispersion of the distributions of participants by GROUP represents the model assumption of homogeneity of variance.

The F ratio and statistical significance

- The *F* ratio is a ratio of weighted (by degrees of freedom) BETWEEN GROUPS VARIATION to WITHIN GROUPS VARIATION. *see note on next page
- The *F* ratio will be large enough to be "statistically significant" if the BETWEEN GROUPS VARIATION "outweighs" the WITHIN GROUPS VARIATION enough, relative to the degrees of freedom.
- That is, if the *F* ratio is equal to or exceeds that which we would expect by chance (e.g., less than 5 times out of 100) if there were no differences between groups (e.g., null hypothesis), then we would reject the null hypothesis and declare a statistically significant difference, i.e., our "model" of experimental grouping explains more of the variation in rhythm perception than we would expect by chance if there were no systematic difference.
- These data do indeed suggest a significant difference according to group… that the variation explained by grouping is beyond that which we'd expect by chance if there were no systematic differences. In fact, the variation explained by the "model" is: [$\mathbf{SS_{between}}$/ $\mathbf{SS_{total}}$] = 4116.28/18759.84 = .22 = 22%

Oneway ANOVA: Sightsinging as a function of experimental condition (N = 150)

Source	*SS*	*df*	*MS*	*F*	*p*
Condition	4116.28	2	2058.14	20.66	<.001
Error/Residual	14643.56	147	99.62		
Total	18759.84	149			



Note: The shift in blue dots along the x-axis is a graphical "jitter" – each cluster of blue dots would have the same "x indicator" which refers to their group.

* The degrees of freedom, i.e., number of objects free to vary, for the BETWEEN GROUPS VARIATION is the number of groups in the independent variable minus 1 (e.g., 3-1 = 2). The degrees of freedom for the WITHIN GROUPS VARIATION is the total degrees of freedom minus the BETWEEN GROUPS degrees of freedom. The total degrees of freedom is the sample size minus 1 (e.g., 150-1 = 149), as such the degrees of freedom for the WITHIN GROUPS VARIATION is 149-2 = 147. Dividing the sums of squares (e.g., the "raw" variation) of the BETWEEN/condition and WITHIN GROUPS/error/residual by their respective degrees of freedom yields the mean square for each. The F statistic is the ratio of the $MS_{Between}/MS_{Within}$.