Math 247: Introduction to Analysis of Variance: One-Way ANOVA (Sections 11.2, 11.3)

A story by Patty Neighmond, reporting for NPR, describes a study investigating the role of music in pain reduction. The study used a random sample of kids who were recovering from surgery. The kids were randomly assigned to 3 groups. Each child either 1) listened to an audio book, 2) listened to music, or 3) sat with noise-cancelling ear phones for 30 minutes. The researchers found that kids in both the music and audio book groups experienced pain reduction levels comparable to over-the-counter pain medication while the control group enjoyed no such benefits.

The researchers used the 10-point FACES scale:

Wong-Baker FACES Pain Rating Scale

Here is a data set that approximates the original findings (each number is a pain level reported by the kids in the study, after the pain intervention method was used):

Audio Book:	5	6	7	2	6	3	4	8	5	4	N = 10 Mean = 5.000	StDev = 1.826
Music:	5	4	4	7	6	4	6	4	3	5	N = 10 Mean = 4.800	StDev. = 1.229
Control:	4	8	7	6	10	6	10	8	5	6	N = 10 Mean = 7.000	StDev. = 2.000

Did any of the methods seem to have helped the kids <u>in the sample groups</u> experience less pain, <u>on average</u>?

How could we determine whether this data provides evidence that either (or both) interventions could help <u>all</u> children, on average, to experience significantly less pain after surgery?

What types of significance are we talking about here?

What are the two main variables in this study?

Which variable is the Treatment (Factor/Explanatory/Independent)?_____

What kind of variable is it?

What are the different categories of this variable? (These are called the "Factor Levels" or just "Levels")

Which variable is the Outcome (Response/Dependent)? _____

What kind of variable is it?

Conduct a hypothesis test to determine whether there is a significant difference in pain levels for the different pain management methods.

(Step 1) _____

(Step 2) _____ Choose significance level:_____

Choose Test: _____

Check conditions, make assumptions.

Conditions for One-Way ANOVA:
1. **Random Samples and Independent Measurements.** Observations can be thought of as a random sample from a population, and individual measurements **within** a group are independent of one another.
2. **Independent groups.** The groups are independent from one another.
3. **Same variance.** The populations must all have equal variances.
4. **Normal Distribution or Large Sample**. The distribution of values is Normal in each group's population OR the sample size is large (at least 25 in each group).

Check the conditions for this study.

Step 2 (continued)

Note: We can check the assumptions of normality and equal variance by graphing and analyzing a boxplot of each data set.
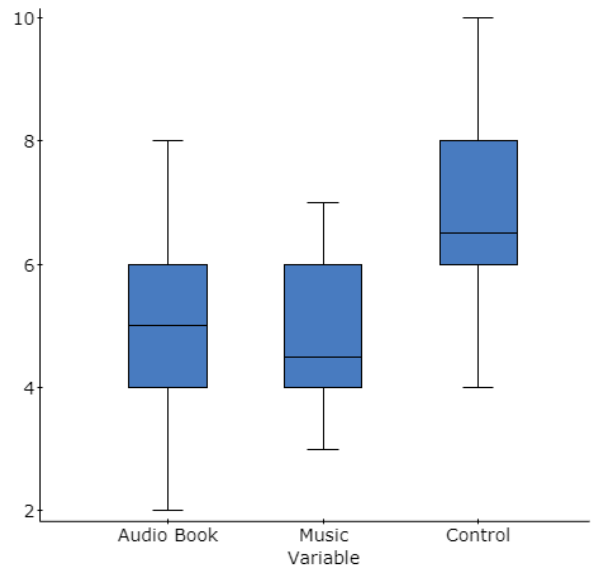
We can also check the assumption of equal variance by examining the standard deviation of each group and applying the "**Two times the Least**" rule. "Two times the Least" means the largest standard deviation should be no more than two times the least standard deviation.

(Step 3) _____

One-Way ANOVA Test on StatCrunch:
Open the StatCrunch file for this data on the WrightMath website.
Use StatCrunch to perform a One-way ANOVA test.
Under "Options", select the "Tukey" box. Click "Okay" and "Okay".

ANOVA table

Source	DF	SS	MS	F-Stat	P-value
Columns	2	29.6	14.8	5.0201005	0.014
Error	27	79.6	2.9481481		
Total	29	109.2			

(Step 4): _____

Follow up: (Post hoc procedures):
Tukey HSD results (95% level)

Audio Book subtracted from

	Difference	Lower	Upper	P-value
Music	-0.2	-2.10	1.70	0.9633
Control	2	0.096	3.90	0.0381

Music subtracted from

	Difference	Lower	Upper	P-value
Control	2.2	0.296	4.10	0.0211

<u>Analyze</u> the results from the Tukey Test.

More on the ANOVA test and the F-statistic:

What does the F-statistic measure? $F = \dfrac{\text{The variation BETWEEN the groups}}{\text{The variation WITHIN the groups}}$

Visual ANOVA: You can use graphs (boxplots, dotplots, histograms) for the groups (factor levels) and compare the difference in means between the groups with the overall variation within the groups. **See pages 541 and 542 in the textbook to see how you can use graphs of the data to measure the different types of variation.**

Returning to the example on pain management in children, we will explore how we measure variation BETWEEN groups and variation WITHIN groups.

Variation **WITHIN** the groups: By hand, calculate the sample standard deviation for the pain levels in the Audio Book group.

Group 1: Audio Book: 5 6 7 2 6 3 4 8 5 4   Mean = 5.000

What is the Sum of the Squared Error (SSE)?

How could you use the sample standard
deviation to work backward to the SSE?

Factor Level (Treatment Group) x = pain level	Error $x - \bar{x}$	Squared Error $(x - \bar{x})^2$	Sum of Squares SSE	"Average" of SSE = the Variance $s^2 = \dfrac{SSE}{n-1}$	Standard Deviation $s = \sqrt{\dfrac{SSE}{n-1}}$
5	0	0	30	3.33333	1.82574
6	1	1			
7	2	4			
2	-3	9			
6	1	1			
3	-2	4			
4	-1	1			
8	3	9			
5	0	0			
4	-1	1			

Now we'll actually find the numerical measurements of the different types of variation (called "ERROR") and summarize it all in an ANOVA table:

Here is the data set and summary statistics for reference:

| | | | | | | | | | | | N = 10 | Mean = 5.0 | s = 1.826 |
Audio Book: 5 6 7 2 6 3 4 8 5 4   N = 10   Mean = 5.0   s = 1.826
Music: 5 4 4 7 6 4 6 4 3 5   N = 10   Mean = 4.8   s = 1.229
Control: 4 8 7 6 10 6 10 8 5 6   N = 10   Mean = 7.0   s = 2.000

First, well find the variation **WITHIN** all the groups (use the standard deviations for each group to do this).

$$SSE_{Within} = \sum SSE_i = \sum (n_i - 1) \cdot s_i^2$$

Next, we'll find the variation **BETWEEN** the groups.

Find the Grand Mean (the average of all the data values or the weighted Mean of the Means)

Grand Mean = _____ $168/30 = 5.6$ or $(5 + 4.8 + 7)/3 = 5.6$

Plot the individual means from each group and also the Grand Mean on the dotplot.

To measure the variation **BETWEEN** the groups, find the Sum of Squares for the means of the groups (compared to the Grand Mean):

$$SSE_{Between} = \sum n_i (\overline{x}_i - \overline{\overline{x}}_{gm})^2$$
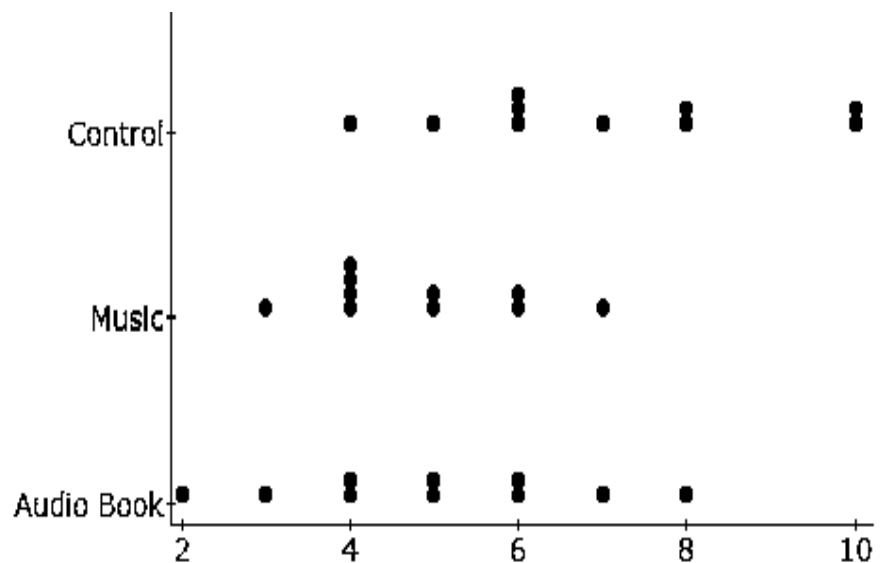
The ANOVA F-table. We'll fill this in by hand using our results from the previous page. Note that this is the same table provided by StatCrunch. This table gives us a concrete way to see how the "signal" (a true difference in the means of the groups) relates to the "noise" (the variation WITHIN the groups)

$$F = \frac{signal}{noise} = \frac{MS_{Between}}{MS_{Within}}$$

Source of Variation	(degrees of freedom) df	(Sum of Squares) SS	(Mean Squares) MS	(F-statistic) $F = \frac{MS_{Between}}{MS_{Within}}$	P-value
BETWEEN Groups Treament/ Factor/ Explained/ SIGNAL					
WITHIN Groups Error/ Residual/ Unexplained/ NOISE					
Total					

F Densities
For Different Values of d1 and d2