One-Way ANOVA with Independent Measures

Suppose we wish to determine whether a set of sample means $(\bar{X}_1, \bar{X}_2, \cdots)$ show a statistically significant difference. Here, rather than the t-tests that compare two-samples, we can use Analysis of Variance (ANOVA). Like the t-test, the test statistic for ANOVA is a ratio that quantifies differences between means in the numerator and variability within groups in the denominator. However, since ANOVA considers more than two groups we can't just take a simple difference in the numerator, instead we use the variability between groups. For the ANOVA this test statistic is called F, and under the null hypothesis F follows an F-distribution. Just like the t-distribution, the shape of the F-distribution depends on the "degrees of freedom." In this case we will need to keep track of both the degrees of freedom for the numerator and degrees of freedom for the denominator, sometimes denoted ν_1 and ν_2. Here is the setup for a one-way ANOVA with independent measures from k groups:

$$F = \frac{SS_{\text{between groups}}/df_{\text{between groups}}}{SS_{\text{within groups}}/df_{\text{within groups}}}$$

$$SS_{\text{between groups}} = \sum_{\text{groups}} n_i(\bar{X}_i - \bar{X}_{\text{All}})^2 \qquad df_{\text{between groups}} = k - 1$$

$$SS_{\text{within groups}} = \sum_{\text{groups}} \sum_{\text{scores}} (X - \bar{X}_i)^2 \qquad df_{\text{within groups}} = N - k$$

where the grand mean $\bar{X}_{All}$ is the mean of all scores (no matter which group) and N is the total number of samples (no matter which group)

$$N = n_1 + n_2 + \cdots + n_k.$$

One of the key ideas in ANOVA is that variability can be "partitioned" into different sources. It turns out that the different sums of squares and the degrees of freedom are related to each other, and you can often simplify the calculations by thinking about these relationships.

$$SS_{\text{total}} = \sum (X - \bar{X}_{\text{All}})^2$$
$$SS_{\text{total}} = SS_{\text{between groups}} + SS_{\text{within groups}}$$
$$df_{\text{total}} = df_{\text{between groups}} + df_{\text{within groups}} = N - 1$$

Finally, just like the previous hypothesis tests, we would like to quantify effect size. For the ANOVA effect size is determined by "Variance Accounted For", which for independent measures, is

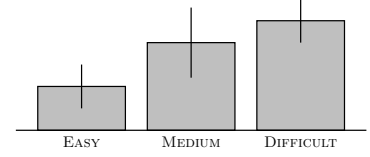$$\eta^2 = \frac{SS_{\text{between groups}}}{SS_{\text{total}}}$$
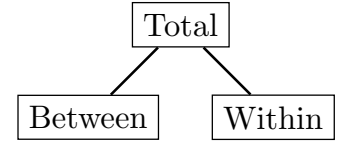

Figure 1: Data from $k = 3$ groups.

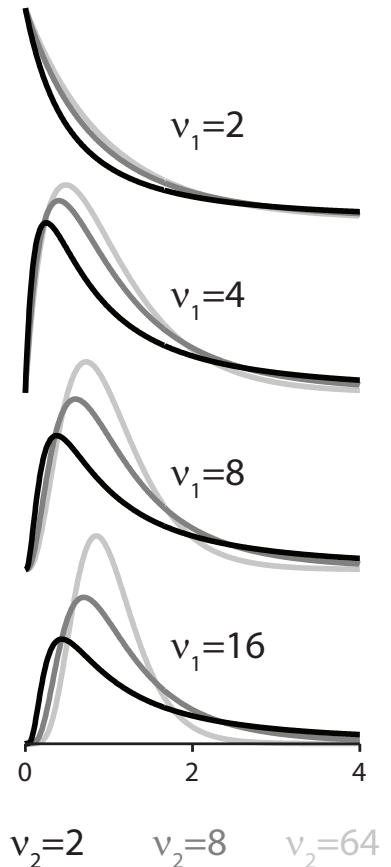Figure 2: Partitioning the Sum of Squares for the One-Way ANOVA with Independent Measures

Figure 3: F-distributions illustrating how the shape varies with different combinations of ν_1 and nu_2. Note that we always have $F \geq 0$, and that $F = 0$ is the mode when $\nu_1 \leq 2$

Example

Suppose you collect data using a post-test only design with 15 participants in three different groups, and observe the data below. What is the value for the test statistic F?

	Group 1	Group 2	Group 3
	4	0	1
	3	1	2
	6	3	2
	3	1	0
	4	0	0
$\bar{X} =$			
$SS =$			

Source	SS	df	MS	F
Between Groups				
Within Groups				
Total				

Solution

	Group 1	Group 2	Group 3
	4	0	1
	3	1	2
	6	3	2
	3	1	0
	4	0	0
$\bar{X} =$	4	1	1
$SS =$	6	6	4

Based on the data we know $k = 3$, and $n_1 = n_2 = n_3 = 5$. Let's start by figuring out $SS_{\text{between groups}}$

$$SS_{\text{between groups}} = \sum_{\text{groups}} n_i(\bar{X}_i - \bar{X}_{\text{All}})^2$$

After calculating the group means (above) we only need the grand mean...

$$\bar{X}_{\text{All}} = (4 + 3 + 6 + \cdots)/15 = 2$$

Now we can fill in

$$SS_{\text{between groups}} = \sum_{\text{groups}} n_i(\bar{X}_i - \bar{X}_{\text{All}})^2 = 5(4-2)^2 + 5(1-2)^2 + 5(1-2)^2 = 30$$

$$df_{\text{between groups}} = k - 1 = 3 - 1 = 2$$

Now let's figure out $SS_{\text{within groups}}$

$$SS_{\text{within groups}} = \sum_{\text{groups}} \sum_{\text{scores}} (X - \bar{X}_i)^2 = \sum_{\text{groups}} SS_i = 6 + 6 + 4 = 16$$

$$df_{\text{within groups}} = N - k = 15 - 3 = 12$$

Plugging these numbers into the equation for *F* and using the partitioning to calculate the totals we can fill in the table

Source	SS	df	MS	F
Between Groups	30	2	15	F=11.25
Within Groups	16	12	1.33	
Total	46	14		

Further Questions

Is this F-ratio statistically significant? What is the effect size?