

## A warmup

Let's start with some short, quiz-like conceptual Qs:

1. What is the purpose of ANOVA? In other words, what type of research questions does ANOVA help us answer?
2. Make up a specific example of a (small) data set on which you can apply an ANOVA test.
3. What is the key theoretical idea that ANOVA is based on? In other words, what statistical attributes of a data set does an ANOVA test try to discern?
4. What conclusion or inference does a "significant" ANOVA test statistic lead to?

## ANOVA recap

A good way to summarize the key ideas and computational steps is via ...

The ANOVA table

variability source	sum of squares	$df$	mean squares	$F$ -statistic
factor (between)	$SS_B$	$k - 1$	$MS_B = \frac{SS_B}{k - 1}$	$F = \frac{MS_B}{MS_W}$
residuals (within)	$SS_W$	$n - k$	$MS_W = \frac{SS_W}{n - k}$	
total	$SS_T = SS_B + SS_W$	$n - 1$		

$k$  = number of groups;  $n$  = total sample size

### Exercise/Example:

Compute the ANOVA table given the following data

	Factor levels			
	good	better	best	stupendous
$n_j$	10	12	8	11
$\bar{y}_j$	37	42	40	36
$s_j^2$	10	10	10	10

$n_j$  = size of group  $j$ ;  $\bar{y}_j$  = mean of group  $j$ ;

$s_j^2$  = variance of group  $j = \frac{\sum_i (y_{ij} - \bar{y}_j)^2}{(n_j - 1)}$ .

### Solution advise

Since you're given the mean and variances, much of the grunt-work of the computations has already been done for you. All that remains is to compute  $SS_B$  and  $SS_W$ , which I want you to figure out and deeply understand how to do.

## ANOVA post-hoc diagnostics

Let's extend the example from last class: A survey is carried out on the number of job offers that a sample of students had at the time of graduation. This time the survey includes 5 majors: Data Science, Biology, Economics, Global Management, and History.

A one-way ANOVA analysis finds evidence of a statistically significant difference in the mean job offers. But, what more might we want to know?

- Is the difference between DS and Biology statistically significant?
- What about between Economics and Global Management?
- Etc?

**An exercise** to help us wrangle with these Qs?

1. Suppose we want to compare the mean values of 3 different groups, in every possible pairwise combination. How many combinations exist?
2. Similarly, suppose we have 4 different groups. How many pairwise combinations exist?
3. What is the general pattern or formula for  $k$  groups?

### **Tukey's HSD test**

That is the technical name of the series of tests for pairwise comparisons of mean values after ANOVA. It works almost the same as a standard  $t$ -test for comparing the means of any two independent samples – the only difference is in the standard error formula.