One-way ANOVA

Overview

ANOVA (Analysis of Variance) tests whether the means of three or more groups are equal. It compares variation *between* groups to variation *within* groups.
Why not multiple t-tests? With k groups, you'd need $\binom{k}{2}$ pairwise t-tests, which inflates the Type I error rate. ANOVA controls the overall error rate.
Hypotheses:

- H_0: $\mu_1 = \mu_2 = \cdots = \mu_k$ (all group means are equal)
- H_a: At least one mean is different

Assumptions:

1. Independence of observations
2. Normality within each group (or large samples)
3. Equal variances across groups

The ANOVA table

Source	Sum of Squares	df	Mean Square	F
Between groups	SSB	$k-1$	$\text{MSB} = \dfrac{\text{SSB}}{k-1}$	$F = \dfrac{\text{MSB}}{\text{MSW}}$
Within groups	SSW	$N-k$	$\text{MSW} = \dfrac{\text{SSW}}{N-k}$	
Total	SST	$N-1$		

where k = number of groups and N = total number of observations.
Key relationship: SST $=$ SSB $+$ SSW

Computing sums of squares

Let $\bar{x}_i$ = mean of group i, n_i = size of group i, $\bar{x}$ = grand mean.

$$\text{SSB} = \sum_i n_i(\bar{x}_i - \bar{x})^2 \qquad \text{SSW} = \sum_i \sum_j (x_{ij} - \bar{x}_i)^2$$

Student Learning Support, Teaching and Learning Centre
studentlearning@ontariotechu.ca
ontariotechu.ca/studentlearning

Decision rule

Reject H_0 if $F > F_{\alpha,k-1,N-k}$ (critical value from F-table), or if p-value $< \alpha$.

Example: Three teaching methods are tested. Is there a difference in mean scores?

Method A	Method B	Method C
78, 82, 75, 80, 79	85, 89, 82, 88, 86	90, 88, 95, 92, 91

Step 1: Calculate means. $\bar{x}_A = 78.8$, $\bar{x}_B = 86$, $\bar{x}_C = 91.2$, Grand mean: $\bar{x} = 85.33$

Step 2: Calculate SSB.

$$\text{SSB} = 5(78.8 - 85.33)^2 + 5(86 - 85.33)^2 + 5(91.2 - 85.33)^2 = 213.2 + 2.2 + 172.3 = 387.7$$

Step 3: Calculate SSW.

Group A: $(78 - 78.8)^2 + (82 - 78.8)^2 + (75 - 78.8)^2 + (80 - 78.8)^2 + (79 - 78.8)^2 = 26.8$

Group B: $(85 - 86)^2 + (89 - 86)^2 + (82 - 86)^2 + (88 - 86)^2 + (86 - 86)^2 = 30$

Group C: $(90 - 91.2)^2 + (88 - 91.2)^2 + (95 - 91.2)^2 + (92 - 91.2)^2 + (91 - 91.2)^2 = 26.8$

SSW $= 26.8 + 30 + 26.8 = 83.6$

Step 4: Build ANOVA table.

Source	**SS**	**df**	**MS**	**F**
Between	387.7	2	193.85	27.83
Within	83.6	12	6.97	
Total	471.3	14		

Step 5: Decision. Critical value: $F_{0.05,2,12} = 3.89$

Since $F = 27.83 > 3.89$, reject H_0. There is a significant difference among the three methods.

Note: ANOVA tells us *at least one* mean differs, but not *which* ones. Post-hoc tests (e.g., Tukey's HSD) identify specific differences.