CEE 260/MIE 273: Probability and Statistics in Civil Engineering

Lecture 6C: Analysis of Variance (ANOVA)

Jimi Oke

UMass Amherst

College of Engineering

November 25, 2025

Outline

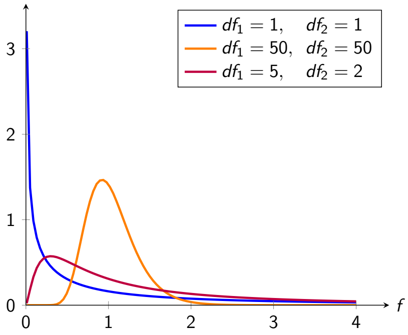
Inferences concerning two population variances

It is sometimes necessary to compare two population variances (or standard deviations) using the sample sum of squares S_1^2 and S_2^2.

For this, the F-test is used. The test statistic is given by:

$$f = \frac{S_1^2/\sigma_1^2}{S_2^2/\sigma_2^2} \tag{1}$$

- The F-distribution has 2 parameters: df_1 and df_2.
- Coined in 1934 by George W. Snedecor in honor of Sir Ronald A. Fisher who developed the ANOVA method, among other achievements.



Analysis of Variance (ANOVA)

So far, we have considered inference for:

- one-sample cases
- two-sample cases

ANOVA is a methodology for comparing means across multiple groups.

Examples of experiments/studies requiring ANOVA

- Effects of five different brands of gasoline on automobile engine operating efficiency (mpg)
- Effects of the presence of four different sugar solutions on bacterial growth
- Investigating if hardwood concentration in pulp has an effect on tensile strength of bags made from the pulp
- Deciding whether the color density of fabric specimens depends on the amount of dye used

Single-factor ANOVA

Comparison of more than two populations or treatment means.

$$
\begin{aligned}
k &= \text{number of populations/treatments} \\
\mu_1 &= \text{mean of population 1 average response when treatment 1 is applied} \\
&\vdots \\
\mu_k &= \text{mean of population } k \text{ average response when treatment } k \text{ is applied}
\end{aligned}
$$

Hypotheses:

$$
\begin{aligned}
H_0 :& \quad \mu_1 = \mu_2 = \cdots = \mu_k \\
H_1 :& \quad \text{at least two of the other } \mu_i\text{'s are different}
\end{aligned}
$$

The ANOVA F-test

Assumption All k populations are normally distributed with means μ_i and equal variance σ^2. The samples are independent and random.

Hypotheses

$$\begin{aligned} H_0 &: \quad \mu_1 = \mu_2 = \cdots = \mu_k \\ H_1 &: \quad \text{at least one the other } \mu_i\text{'s differs from the others} \end{aligned}$$

Test Statistic

$$f = \frac{\text{mean square among groups}}{\text{mean square error}}$$

Compare (a) Find critical value $F_{(1-\alpha, df_1, df_2)}$.
If $f \leq F_{(1-\alpha, df_1, df_2)}$, **fail to reject H_0** *OR*
(b) Find p-value.
If $\alpha \leq p$-value, **fail to reject H_0**

Different sums of squares in ANOVA

Total sum of squares (SST)

$$SST = \sum_{i=1}^{k} \sum_{j=1}^{n_i} (x_{ij} - \overline{x})^2 \tag{2}$$

where:

- x_{ij} is the observed quantity j in group i
- n_i is the number of observations (sample size) in group i
- $\overline{x}$ is the **grand mean** from all k samples:

$$\overline{x} = \frac{1}{n} \sum_{i=1}^{k} \sum_{j=1}^{n_i} x_{ij} \tag{3}$$

and $n = n_1 + n_2 + \cdots + n_k$

Different sums of squares in ANOVA (cont.)

Sum of squares between groups (SSG)

Measures the variance *among* the k sample means

$$SSG = \sum_{i=1}^{k} n_i(\overline{x_i} - \overline{x})^2 \tag{4}$$

(Also called **sum of squares for treatments** ($SSTr$))

Sum of squares for error (SSE)

Measures the pooled variance *within* each of the k samples

$$SSE = (n_1 - 1)s_1^2 + (n_2 - 1)s_2^2 + \cdots + (n_k - 1)s_k^2 = \sum_{i=1}^{k}(n_i - 1)s_i^2 \tag{5}$$

where s_i^2 is the sample variance for group i.

Fundamental identities and d.o.f.

Fundamental identity

$$SST = SSG + SSE \tag{6}$$

(You can prove this identity as an exercise.)

Degrees of freedom (df):

$$\begin{aligned} df(SST) &= n-1 & (7)\\ df(SSG) &= k-1 & (8)\\ df(SSE) &= n-k & (9)\end{aligned}$$

It also follows that:

$$df(SST) = df(SSG) + df(SSE) \tag{10}$$

Variances among and within groups

Ultimately, we want to determine how the variance is porportioned among the samples.

Mean square among groups (MSG)

$$MSG = \frac{1}{k-1} \times SSG = \frac{1}{k-1} \sum n_i(\overline{x_i} - \overline{x})^2 \tag{11}$$

- Variance among the groups/treatments
- Also called **mean square for treatments** (MSTr)

Mean square [for] error (MSE)

$$MSE = \frac{1}{n-k} \times SSE = \frac{1}{n-k} \sum_{i=1}^{k} (n_i - 1)s_i^2 \tag{12}$$

- Pooled variance within each of the groups.

Summary of ANOVA notation and sums

$$
\begin{aligned}
k &= \text{number of groups/samples/treatments/populations} \\
n &= n_1 + n_2 + \cdots + n_k \\
\overline{x} &= \frac{1}{n}\sum x_{ij} \quad \text{(grand mean)} \\
SST &= \sum (x_{ij} - \overline{x})^2 = \sum x_{ij}^2 - CM \\
SSG &= \sum_{i=1}^{k} n_i(\overline{x_i} - \overline{x})^2 = \sum \frac{T_i^2}{n_i} - CM \\
T_i &= \sum_{j} x_{ij} \quad \text{(total of the observations for group } i\text{)} \\
CM &= \frac{1}{n}\left(\sum x_{ij}\right)^2 \quad \text{(correction for the mean)} \\
SSE &= \sum_{i=1}^{k} (n_i - 1)s_i^2 = SST - SSG \\
MSG &= \frac{1}{k-1} SSG \qquad MSE = \frac{1}{n-k} SSE
\end{aligned}
$$

Test statistic

The test procedure compares a measure of differences among the sample means ("between-samples" variation) to a measure of variation calculated from within each of the samples.

Test statistic for one-way ANOVA

$$F = \frac{MSG}{MSE} \tag{13}$$

- $H_0 : \mu_1 = \mu_2 = \cdots = \mu_k$
- MSE is an unbiased estimate of σ^2 whether or not H_0 is true
- MSG is an unbiased estimate of σ^2 ONLY when H_0 is true
- When H_0 is not acceptable, $\mathbb{E}(MSG) > \mathbb{E}(MSE) = \sigma^2 \implies f > 1$
- When H_0 is acceptable, f follows the F-distribution with:

$$df_1 = k - 1 \tag{14}$$
$$df_2 = n - k \tag{15}$$

The ANOVA F-test

Assumption All k populations are normally distributed with means μ_i and equal variance σ^2. The samples are independent and random.

Hypotheses

$$H_0: \quad \mu_1 = \mu_2 = \cdots = \mu_k$$
$$H_1: \quad \text{at least one the other } \mu_i\text{'s differs from the others}$$

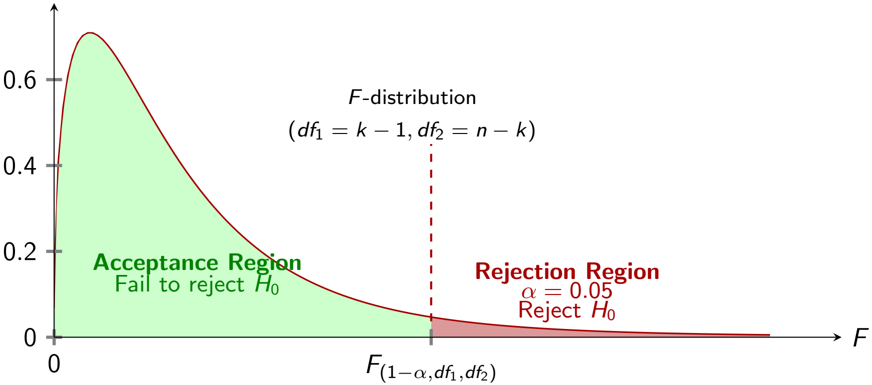
Test Statistic

$$f = \frac{MSG}{MSE}$$

Compare (a) Find critical value $F_{(1-\alpha, df_1, df_2)}$[1] (also called f^*).
If $f \leq F_{(1-\alpha, df_1, df_2)}$, **fail to reject H_0**
If $f > F_{(1-\alpha, df_1, df_2)}$, **reject H_0** *OR*
(b) Find p-value.
If p-value $\geq \alpha$, **fail to reject H_0**
If p-value $< \alpha$, **reject H_0**

[1]Use `f.ppf(1 - alpha, df1, df2)` in Python

ANOVA F-test visualization



Decision Rule

- If $f \le F_{(1-\alpha, df_1, df_2)}$: Fail to reject H_0 (means are equal)
- If $f > F_{(1-\alpha, df_1, df_2)}$: Reject H_0 (at least one mean differs)

ANOVA table

When conducting one-way ANOVA, computations are usually summarized in an ANOVA table

Source of variation	**d.o.f.**	**Sum of Squares**	**Mean Square**	f
Groups	$k-1$	SSG	$MSG = SSG/(k-1)$	MSG/MSE
Error	$n-k$	SSE	$MSE = SSE/(n-k)$	
Total	$n-1$	SST		

Using the ANOVA table

Example 1: Elastic moduli of alloys

An experiment was performed to measure the elastic modulus (GPa) of an alloy produced using three different casting processes. Let μ_1, μ_2, μ_3 denote the true average elastic moduli for the 3 different processes. Using the ANOVA table given below, test the null hypothesis that all three means are equal (using the *p*-value approach).

Source of variation	d.o.f.	Sum of Squares	Mean Square	*f*
Treatments	2	7.93	3.965	?
Error	19	6.00	.3158	
Total	21	13.93		

Using the ANOVA table

Example 1: Elastic moduli of alloys (cont.)

Step 1. First, we compute f:

$$f = \frac{MSG}{MSE} = \frac{3.965}{.3158} = 12.56$$

Step 2. Second, we find the p-value[a]:

$$p\text{-value} \quad = \quad 1 - F(12.56, 2, 19) = 0.0003$$

Step 3. We conclude: since the p-value is very small, we can **reject the null hypothesis** at any reasonable significance level.

[a] In MATLAB, use: `1 - f.cdf(12.56, 2, 19)` OR `f.sf(12.56, 2, 19)`

One-way ANOVA in practice

Example 2: Recording tape quality

In a effort to improve the quality of recording tapes, the effects of four kinds of coatings A, B,C, D on the reproducing quality of sound are compared.
Independent samples are obtained foreach kind of coating.
The following values on distortion are measured:

Coating	**Observations**	**Sample Sizes**
A	10, 15, 8, 12, 15	5
B	14, 18, 21, 15	4
C	17, 16, 14, 15, 17, 15, 18	7
D	12, 15, 17, 15, 16, 15	6

With the help of such a sample we want to decide if the four different coatings result in different mean distortions, i.e. test if $\mu_A = \mu_B = \mu_C = \mu_D$ (use $\alpha = 0.05$).

One-way ANOVA in practice

Example 2: Recording tape quality (cont.)

Step 1. Assemble the parameters:

$$
\begin{aligned}
k &= 4 \\
n_1 &= 5, n_2 = 4, n_3 = 7, n_4 = 6 \\
T_1 &= 60, T_2 = 68, T_3 = 112, T_4 = 90 \\
n &= \sum n_i = 22 \\
\sum x_{ij} &= \sum T_i = 330
\end{aligned}
$$

One-way ANOVA in practice

Example 2: Recording tape quality (cont.)

Step 2. Compute the sums:

$$
\begin{aligned}
CM &= \frac{1}{n}\left(\sum x_{ij}\right)^2 = \frac{330^2}{22} = 4950 \\
SST &= \sum x_{ij}^2 - CM = (10^2 + 15^2 + 8^2 + \cdots + 15^2 + 5^2) - CM \\
&= 5112 - 4950 = 162 \\
SSG &= \sum \frac{T_i^2}{n_i} - CM \\
&= \left(\frac{60^2}{5} + \frac{68^2}{4} + \frac{112^2}{7} + \frac{90^2}{6}\right) - CM = 68 \\
SSE &= SST - SSG = 162 - 68 = 94
\end{aligned}
$$

One-way ANOVA in practice

Example 2: Recording tape quality (cont.)

Step 3. Create the ANOVA table:

Source	**d.o.f.**	***SS***	***MS***	***f***
Coating	3	68	$MSG = 22.67$	$MSG/MSE = 4.343$
Error	18	94	$MSE = 5.22$	
Total	21	162		

One-way ANOVA in practice

Example 2: Recording tape quality (cont.)

Step 4. Find critical value:

$$F_{0.95,3,18} = F^{-1}(0.95, 3, 18) = 3.1599 \quad \text{(from Python)}$$

Notes:

- Use `f.ppf(0.95, 3, 18)` to find the critical value.
- The F curve is NOT symmetric. So, `-f.ppf(0.05, 3, 18)` will NOT give the same answer.
- However, the F-curve has the property

$$F_{(\alpha, df_1, df_2)} = \frac{1}{F_{(1-\alpha, df_2, df_1)}} \tag{16}$$

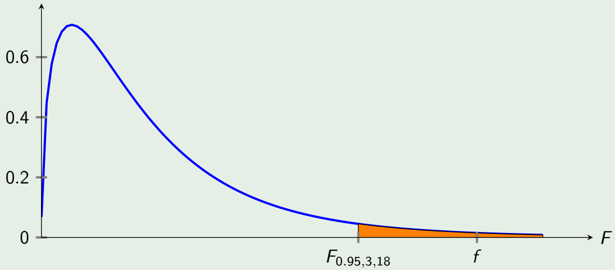
- So, in this case, `f.ppf(0.05, 18, 3)` will give the same answer as `f.ppf(0.95, 3, 18)`.

One-way ANOVA in practice

Example 2: Recording tape quality (cont.)

Step 5. Compare:

$$f = 4.343 > F_{0.95,3,18} = 3.1599$$



Thus, f lies in the critical region, and we **reject the null hypothesis** that the mean distortions from the four coatings are equal.

ANOVA in MATLAB/Python

- The computations in ANOVA are tedious and prone to human error.
- Usually ANOVA is conducted using software platforms such as MATLAB or Python.

Example 3: Investigating fuel efficiency of cars manufactured in different countries

The `carbig` dataset in MATLAB contains variables[a] for 400 cars from the 70s and 80s. The `carsmall` dataset is a subset of `carbig` (years '70, '76 and '82). We want to check if the true average fuel efficiency (MPG) of vehicles from different countries (USA/Europe/Japan) are equal.

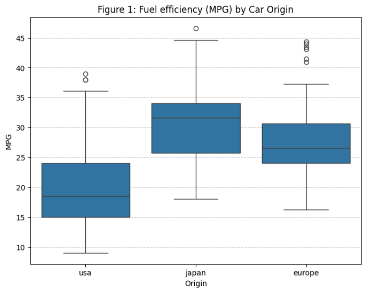
[a] For more on visualizing some of these variables, visit
https://www.mathworks.com/help/stats/examples/visualizing-multivariate-data.html

ANOVA in MATLAB (cont.)

Example 3: Investigating fuel efficiency of cars manufactured in different countries (cont.)

Step 1. Examine the contents of the `6C-anova-cars.ipynb` script. Run the script.

Step 2. Figure 1 is a boxplot of the MPG across the different origins. What can you say about the variance *within* each of the countries/regions, and *among* them?



ANOVA in MATLAB (cont.)

Example 3: Investigating fuel efficiency of cars manufactured in different countries (cont.)

Step 3. The function `scipy.stats.f_oneway` generates the ANOVA results (though not formatted as nicely as in MATLAB—see Appendix). More detail is provided by `statsmodels.stats.oneway.anova_oneway`.

Step 4. What is the p-value for this test? Based on this, do we **reject/fail to reject** the null hypothesis at any reasonable significance level?

Step 5. In Figure 3, we have a notched boxplot, which denotes the confidence interval of the median. What are your observations?

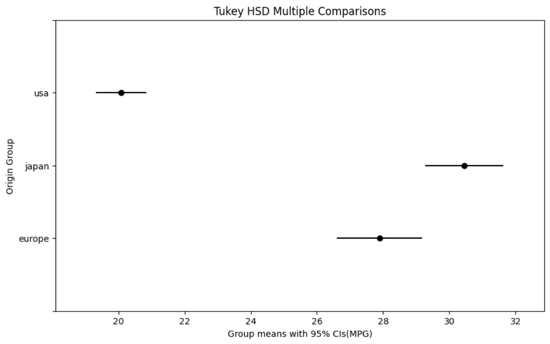
Multiple comparisons in ANOVA

- If after performing a one-way ANOVA, we reject the null hypothesis, naturally we might want to find out which populations significantly differ from each other
- Various methods have been developed for multiple pairwise comparisons (out of scope of this lecture):
  - One option: perform pairwise t-tests between group means using the Bonferroni correction for α
  - Use Tukey's range test, which performs a significance test based on the studentized range distribution

Multiple comparisons in ANOVA (cont.)

Example 3: Investigating fuel efficiency of cars (cont.)

Step 5. Use the `pairwise_tukeyhsd` function from `statsmodels.stats.multicomp` to perform multiple comparisons.



Which regions have significantly different average fuel efficiency?

Summary

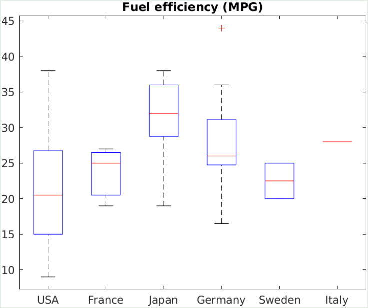
- The analysis of variance (ANOVA) test is used to determine whether the population means across multiple samples are equal
- It uses the test statistic ($f = MSG/MSE$) which is F-distributed
- Reading: Section 7.5 (Open Intro Statistics)
- If interested, read Section 7.4 on "power" analysis (but we will consider this out of scope for the course)

ANOVA in MATLAB (cont.)

Example 3: Investigating fuel efficiency of cars manufactured in different countries (cont.)

Step 1. Examine the contents of the `example3.m` script. Run the script.

Step 2. Figure 1 is a boxplot of the MPG across the different origins (6 countries). What can you say about the variance *within* each of the countries, and *among* them?



ANOVA in MATLAB (cont.)

Example 3: Investigating fuel efficiency of cars manufactured in different countries (cont.)

Step 3. The function `anova1` generates 2 figures: the first is the ANOVA table (Figure 2) and the second a notched boxplot (Figure 3)

ANOVA Table

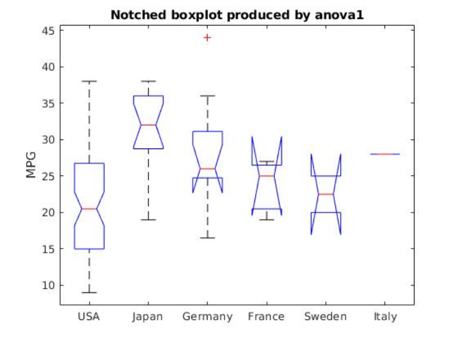
Source	SS	df	MS	F	Prob>F
Groups	1629.87	5	325.974	6.56	3.14631e-05
Error	4375.41	88	49.721		
Total	6005.28	93			

Step 4. The column `Prob>F` is the *p*-value. Based on this, we **reject** the null hypothesis at any reasonable significance level.

ANOVA in MATLAB (cont.)

Example 3: Investigating fuel efficiency of cars manufactured in different countries (cont.)

Step 5. In Figure 3, we have a notched boxplot, which denotes the confidence interval of the median. What are your observations?



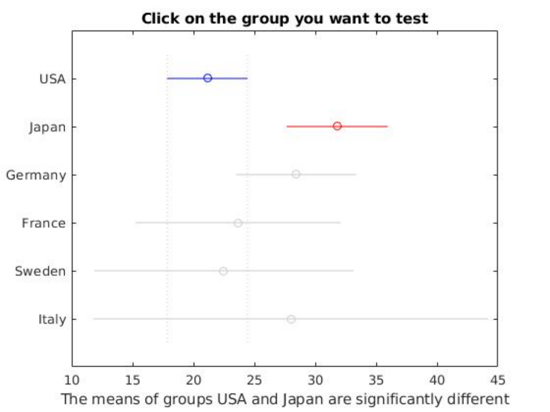
Multiple comparisons in ANOVA

- If after performing a one-way ANOVA, we reject the null hypothesis, naturally we might want to find out which populations significantly differ from each
- Various methods have been developed for multiple pairwise comparisons (out of scope of this lecture):
  - One option: perform pairwise t-tests between group means using the Bonferroni correction for α
  - Use Tukey's range test, which performs a significance test based on the studentized range distribution

Multiple comparisons in ANOVA (cont.)

Example 3: Investigating fuel efficiency of cars (cont.)

Step 5. Uncomment lines 13 and 14 and rerun the `example3.m` script. Interact with Figure 5.



Which countries have significantly different average fuel efficiency?

Temporary page!

LaTeX was unable to guess the total number of pages correctly. As the
unprocessed data that should have been added to the final page this e
has been added to receive it.

If you rerun the document (without altering it) this surplus page will g
because LaTeX now knows how many pages to expect for this documen