

Tutorial Note 1: ANOVA

provided by Nuen-tsang YANG

Anova:

- Compare two or more means.
- Null Hypothesis $H_0: \mu_1 = \mu_2 = \dots = \mu_k$
- In order to test the hypothesis, we need to construct an ANOVA table.

Case 1: Raw data

Now, you are provided a raw data. You are asked to perform the F-test.

| | | | | | | |
|----------|----|-----|-----|----|----|----|
| <i>A</i> | 24 | 32 | 67 | 50 | 33 | |
| <i>B</i> | 90 | 108 | 271 | 82 | | |
| <i>C</i> | 15 | 2 | 5 | 14 | 32 | 22 |

Using your calculator:

- *Enter the data to L_1 :*
STAT → EDIT → 24 → ENTER → 32 → ENTER ...
- *Do it again for L_2, L_3 :*
...
- *ANOVA test:*
STAT → ► → ► (*choose ANOVA test*) → ENTER → 2nd → L_1 → , → 2nd → L_2 → , → 2nd → L_3 → (*you will see ANOVA(L_1, L_2, L_3)*)ENTER

Using the result from your calculator, you can fill the ANOVA table and get p-value for the test,

| <i>Source</i> | <i>df</i> | <i>SS</i> | <i>MS</i> | <i>F</i> |
|---------------------------------|-----------|-----------|-----------|----------|
| <i>Between(orFactororGroup)</i> | 2 | 37910.18 | 18955.09 | 8.8069 |
| <i>Within(orError)</i> | 12 | 25827.55 | 2152.30 | |
| <i>Total</i> | 14 | 63737.73 | | |

So, the p-value is $p = 0.004427 < 0.05 \rightarrow \text{rej. } H_0$.

Case 2: Given summary statistics of the groups

If you are provided the summary statistics of the groups rather than the raw data, you may want to work with the formula in your anova lecture. The example for this case is the problem 4 in your sample midterm 3.

$$\begin{aligned}n_A &= 5 & \bar{Y}_A &= 63.35 & S_A &= 3.47 \\n_B &= 5 & \bar{Y}_B &= 69.3 & S_B &= 2.93 \\n_C &= 4 & \bar{Y}_C &= 80.35 & S_C &= 2.71 \\n_D &= 5 & \bar{Y}_D &= 66.24 & S_D &= 2.9\end{aligned}$$

$$\begin{aligned}df_B &= k - 1 = 4 - 1 = 3 \\df_W &= N - k = 19 - 4 = 15\end{aligned}$$

$$\begin{aligned}\bar{Y} &= \frac{5 \times \bar{Y}_A + 5 \times \bar{Y}_B + 4 \times \bar{Y}_C + 5 \times \bar{Y}_D}{5 + 5 + 4 + 5} \\&= 69.2553 \\SSB &= 5 \times (\bar{Y}_A - \bar{Y})^2 + 5 \times (\bar{Y}_B - \bar{Y})^2 + 4 \times (\bar{Y}_C - \bar{Y})^2 + 5 \times (\bar{Y}_D - \bar{Y})^2 \\&= 712.2 \\SSW &= (5 - 1) \times S_A^2 + (5 - 1) \times S_B^2 + (4 - 1) \times S_C^2 + (5 - 1) \times S_D^2 \\&= 139.6 \\MSB &= \frac{SSB}{df_B} = \frac{712.2}{3} = 237.4 \\MSW &= \frac{SSW}{df_W} = \frac{139.6}{15} = 9.31 \\F &= \frac{MSB}{MSW} = \frac{237.4}{9.31} = 25.58\end{aligned}$$

Now, you can fill the ANOVA table.

| Source | df | SS | MS | F |
|--------------------------|----|-------|-------|-------|
| Between(orFactororGroup) | 3 | 712.2 | 237.4 | 25.58 |
| Within(orError) | 15 | 139.6 | 9.28 | |
| Total | 18 | 851.8 | | |

Next, you have to find the p-value in order to test the hypothesis. Now, using your calculator again:

2nd → DISTR → Fcdf → 25.58 → , → 1 → EE → 99 → , → 3 → , → 15
→ ENTER

Then, you can get the p-value $p = 3.788E^{-6} < 0.001$. So, the null hypothesis is rejected.