

Analysis Of Variances (ANOVA)

Basic Steps

Let's say we collected results of final exam in Winter 2016 semester taken in MATH106, MATH107 and STAT200 courses.

5 classes were selected for each course.

Each class had 30 students.

Table below shows mean score for each class, out of 100 points.

| MATH106 | MATH107 | STAT200 |
|---------|---------|---------|
| 75 | 72 | 82 |
| 74 | 76 | 79 |
| 77 | 81 | 75 |
| 68 | 79 | 78 |
| 78 | 77 | 80 |

Null Hypothesis: there is no significant difference between averages score on the final exam in MATH106, MATH107 and STAT200.

$H_0: \mu_1 = \mu_2 = \mu_3$

ANOVA test examine this statement by analyzing VARIANCES in each category.

Step 1: Compute variance in each group.

The best and easy way is apply Excel function VAR.S() or you can use online variance calculator <http://www.alcula.com/calculators/statistics/variance/>
In our case we will have:

| Variances of means in each category | | |
|-------------------------------------|---------|---------|
| MATH106 | MATH107 | STAT200 |
| 15.3 | 11.5 | 6.7 |

Step 2: Consider all sample variances as a new set of data, let's call it set of variances, and compute mean for this numbers.

$$(15.3 + 11.5 + 6.7)/3 = 11.2$$

This is in Lane textbook terminology Mean Square Error (MSE).

$$MSE = 11.2$$

Step 2: Compute Variance for set of numbers that represent each sample variance. Instead of subtracting 11.2 from each variance (15.3, 11.5, 6.7), then square each result and so on, apply Excel function VAR.S() to column of three numbers: 15.3, 11.5, 6.7. Result will be: 18.6.

Multiply result by number of classes/samples in each category:
 $18.6 \times 5 = 93$

This is in Lane textbook terminology Mean Square Between (MSB).
MSB = 93

Step 3: Compute F-value.

$$\text{F-value} = \text{MSB}/\text{MSE} = 93/11.2 = 8.3$$
$$\text{F-value} = 8.3$$

Step 4: Define degree of freedom df1 and df2.
Let's k stands for number of groups/categories,
and n stands for number of samples in each category.
We have: k = 3 and n = 5.
 $df1 = k - 1 = 3 - 1 = 2$
 $df2 = k(n - 1) = 3(5-1) = 12$

Step 5: Use Excel Statistics function F.DIST.RT(F-value, df1, df2)
to determine P-value. It will give you the area of Right Tail (RT)
in ANOVA F-Distribution.

$$P\text{-value} = \text{F.DIST.RT}(8.3, 2, 12) = 0.00546$$

Step 6: Compare P-value to the significance level.
Significance level is always given and usually is 0.01, 0.05 or 0.10.
The rules are:
if P-value is less than given significance level, Reject Null Hypothesis;
if P-value is greater than given significance level, Do Not Reject Null Hypothesis.

Let's say given significance level is 0.05.
Because $0.00546 < 0.05$ we Reject H_0 .
In other words, ANOVA test confirms that there is a significant difference
between averages of final exam scores in three selected courses.