

Module 8 Analysis of Variance(ANOVA)

1. A traffic engineering study on traffic delay was conducted at intersections with signals on urban streets. Three types of traffic signals were utilized in the study: (1) pretimed, (2) semi-actuated, and (3) fully actuated. Five intersections were used for each type of signal. The measure of traffic delay used in the study was the average stopped time per vehicle at each of the intersections (seconds/vehicle). The data follow.

Pretimed	Semi-actuated	Fully actuated
36.6	17.5	15.0
39.2	20.6	10.4
30.4	18.7	18.9
37.1	25.7	10.5
34.1	22.0	15.2

Compute the ANOVA table and test the hypothesis of no difference among the mean traffic delays of the signal types with the suitable test at the 0.05 level of significance.

Solution: $\bar{x}_{1.} = \frac{36.6+39.2+30.4+37.1+34.1}{5} = 35.48$, $\bar{x}_{2.} = \frac{17.5+20.6+18.7+25.7+22.0}{5} = 20.9$, $\bar{x}_{3.} = \frac{15.0+10.4+18.9+10.5+14}{5} = 14$ and $\bar{x}_{..} = \frac{35.48+20.9+14}{3} = 23.46$.

$$\begin{aligned} \text{SS}(\text{total}) &= \sum_{i=1}^3 \sum_{j=1}^5 (x_{ij} - \bar{x}_{..})^2 = (36.6 - 23.46)^2 + (39.2 - 23.46)^2 + \dots \\ &\quad + (10.5 - 23.46)^2 + (15.2 - 23.46)^2 = 1340.456 \end{aligned}$$

$$\begin{aligned} \text{SS}(\text{within}) &= \sum_{i=1}^3 \sum_{j=1}^5 (x_{ij} - \bar{x}_{i.})^2 = (36.6 - 35.48)^2 + \dots + (34.1 - 35.48)^2 \\ &\quad + (17.5 - 20.9)^2 + \dots + (22.0 - 20.9)^2 + (15.0 - 14)^2 + \dots + (15.2 - 14)^2 \\ &= 137.828 \end{aligned}$$

$$\begin{aligned} \text{SS}(\text{between}) &= \sum_{i=1}^3 5 \times (\bar{x}_{i.} - \bar{x}_{..})^2 = 5 \times ((35.48 - 23.46)^2 + (20.9 - 23.46)^2 + (14 - 23.46)^2) \\ &= 1202.628 \end{aligned}$$

ANOVA table:

Source	Sum of Squares	Degree of Freedom	Mean Square	F value
Between	1202.628	2	601.314	52.35
Within	137.828	12	11.486	
Total	1340.456	14		

The F statistic has 2 and 14 degree of freedoms and a value of 52.35. p-value is close to 0. Alternatively, $F_{2,12}(0.95) = 3.885 < 52.35$, so we reject the null hypothesis. We may conclude that the mean traffic delays of the signal types are significantly different.

2. A school district uses four different methods of teaching their students how to read and wants to find out if there is any significant difference between the reading scores achieved using the four methods. It creates a sample of 8 students for each of the four methods. The reading scores achieved by the participants in each group are as follows:

Method 1	Method 2	Method 3	Method 4
51	82	79	85
87	91	84	80
50	92	74	65
48	80	98	71
79	52	63	67
61	79	83	51
53	73	85	63
54	74	58	93

Use Excel to compute the ANOVA table.

Solution:

Anova: Single Factor						
SUMMARY						
<i>Groups</i>	<i>Count</i>	<i>Sum</i>	<i>Average</i>	<i>Variance</i>		
Column 1	8	483	60.375	214.267857		
Column 2	8	623	77.875	157.553571		
Column 3	8	624	78	164.571429		
Column 4	8	575	71.875	181.553571		
ANOVA						
<i>Source of Variation</i>	<i>SS</i>	<i>df</i>	<i>MS</i>	<i>F</i>	<i>P-value</i>	<i>F crit</i>
Between Groups	1645.34375	3	548.447917	3.05564814	0.0446626	2.94668527
Within Groups	5025.625	28	179.486607			
Total	6670.96875	31				

This time the $p\text{-value} = 0.04466 < 0.05$, and so we reject the null hypothesis, and conclude that there are significant differences between the methods.

3. An experiment was conducted to evaluate the effect of different detergents and water temperatures on the cleanliness of ceramic substrates. The experimenter selected three different detergents based on their pH levels, and conducted a series of experiments at four different water temperatures. Cleanliness was quantified by measuring the contamination of a distilled water beaker after rinsing the parts cleaned using each treatment combination. The coded data are shown in Table. Using Excel to find the ANOVA table.

	DETERGENT A	DETERGENT B	DETERGENT C
Cold	15	18	10
Cool	12	14	9
Warm	10	18	7
Hot	6	12	5

Solution:

ANOVA						
<i>Source of Variation</i>	<i>SS</i>	<i>df</i>	<i>MS</i>	<i>F</i>	<i>P-value</i>	<i>F crit</i>
Rows	68	3	22.66666667	8.242424242	0.015043179	4.757062663
Columns	122.1666667	2	61.08333333	22.21212121	0.001684751	5.14325285
Error	16.5	6	2.75			
Total	206.6666667	11				

Any P-value less than 0.05 would indicate a significant effect. The ANOVA table indicates that there are significant differences between the different detergents and the different water temperatures.

4. A physiologist was interested in learning whether smoking history and different types of stress tests influence the timing of a subject's maximum oxygen uptake, as measured in minutes. The researcher classified a subject's smoking history as either heavy smoking, moderate smoking, or non-smoking. He was interested in seeing the effects of three different types of stress tests a test performed on a bicycle, a test on a treadmill, and a test on steps. The physiologist recruited 9 non-smokers, 9 moderate smokers, and 9 heavy smokers to participate in his experiment, for a total of $n = 27$ subjects. He then randomly assigned each of his recruited subjects to undergo one of the three types of stress test. Here are his resulting data:

	Test		
Smoking History	Bicycle (1)	Treadmill (2)	Step Test (3)
Nonsmoker (1)	12.8	16.2	22.6
	13.5	18.1	19.3
	11.2	17.8	18.9
Moderate (2)	10.9	15.5	20.1
	11.1	13.8	21
	9.8	16.2	15.9
Heavy (3)	8.7	14.7	16.2
	9.2	13.2	16.1
	7.5	8.1	17.8

- Is there sufficient evidence at the $\alpha = 0.05$ level to conclude that smoking history has an effect on the time to maximum oxygen uptake?
- Is there sufficient evidence at the $\alpha = 0.05$ level to conclude that the type of stress test has an effect on the time to maximum oxygen uptake?
- Is there evidence of an interaction between smoking history and the type of stress test?

Solution:

ANOVA						
Source of Variation	SS	df	MS	F	P-value	F crit
Sample	84.8985185	2	42.4492593	12.896703	0.00033479	3.55455715
Columns	298.071852	2	149.035926	45.2792843	9.4727E-08	3.55455715
Interaction	2.81481481	4	0.7037037	0.21379543	0.92734124	2.92774417
Within	59.2466667	18	3.29148148			
Total	445.031852	26				

- (a) The P-value is very small (< 0.001). We reject the null hypothesis that the smoking history parameters are all zero. That is, there is sufficient evidence at the 0.05 level to conclude that smoking history has an effect on the timing of maximum oxygen uptake.
- (b) The P-value is very small (< 0.001). We reject the null hypothesis that the stress test parameters are all zero. That is, there is sufficient evidence at the 0.05 level to conclude that the type of stress test has an effect on the timing of maximum oxygen uptake.
- (c) The P-value, 0.927, is very large. We do not reject the null hypothesis that the interaction terms are all zero. That is, there is insufficient evidence at the 0.05 level to conclude that there is an interaction between smoking history and the type of stress test.