Special Chapter

Introduction to Analysis of Variance (ANOVA)

Analysis of variance (ANOVA) is used to test the hypothesis that three or more population means are equal.

Hypothesis:

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

H_a: Not all population means are equal

Dependent Variable (Response Variable) = Variable to be compared

Independent Variable (Factor or Treatment) = Names of categories or divisions

For example,

(Note)

Assumptions of ANOVA

1. For each population, the response variable is normally distributed.
2. The variance of the response variable is the same for all of the population.
3. The observations must be independent.

A Conceptual Overview

Suppose there are 3 populations. If the means for the three populations are equal, we would expect the three sample means to be close together. Hence, if the variability among the sample means is small, it supports H_0. If the variability among the sample means is large, it supports H_a.

We can then measure these variability by analyzing 2 types of variance estimates:

- Between-Treatments Estimate of Population Variance
- Within-Treatments Estimate of Population Variance

Example:

We are trying to measure how employees at plants located in Atlanta, Dallas, and Seattle understand the concept of Total Quality Management. By doing so, 6 employees from each plant are selected to do the exam and the score for each employee was calculated. To see the difference in abilities of employees in different plants, we can employ ANOVA to test whether there is a difference in mean scores among employees from those plants.

Observation	**Atlanta Plant**	**Dallas Plant**	**Seattle Plant**
1	85	71	59
2	75	75	64
3	82	73	62
4	76	74	69
5	71	69	75
6	85	82	67
Sample Mean	79	74	66
Sample Variance	34	20	32
Sample SD	5.83	4.47	5.66

(Note: Demonstration via picture)

Analysis of Variance (ANOVA)

Test Statistic:

$$F = \frac{MSTR}{MSE}$$

Between-Treatments Estimate of Population Variance

MSTR = Mean Square due to Treatment

$$MSTR = \frac{\sum_{j=1}^{k} n_j(\bar{x}_j - \bar{\bar{x}})^2}{k-1}$$

Where, k – 1 = df between treatment (number of group – 1)

Within-Treatments Estimate of Population Variance

MSE = Mean Square due to Error

$$MSE = \frac{\sum_{j=1}^{k}(n_j - 1)s_j^2}{n_T - k}$$

Where, n_T - k = df among errors (total number of observations – number of group)

Use the information from the previous table to determine whether there is a significant difference among the mean scores of employees from the three locations, given 0.05 level of significance.

Ex 1

Five observations were selected from each of three populations. The data obtained follow.

Observation	Sample 1	Sample 2	Sample 3
1	32	44	33
2	30	43	36
3	30	44	35
4	26	46	36
5	32	48	40
Sample mean	30	45	36
Sample variance	6	4	6.5

a) Compute the between-treatments estimate of variance.
b) Compute the within-treatments estimate of variance.
c) At the 0.05 level of significance, can we reject the null hypothesis that the means of the three populations are equal?

Ex2

Four observations were selected from each of three populations (test scores of KKUIC students from CA, IA, and TM). The data obtained follow.

Observation	**CA**	**IA**	**TM**
1	165	174	169
2	149	164	154
3	156	180	161
4	142	158	148
Sample mean	153	169	158
Sample variance	96.67	97.33	82

a) Compute the between-treatments estimate of variance.
b) Compute the within-treatments estimate of variance.
c) At the 0.05 level of significance, can we reject the null hypothesis that the means of the three populations are equal?

F - Distribution ($\alpha = 0.05$ in the Right Tail)

Denominator Degrees of Freedom df_2 \ Numerator Degrees of Freedom df_1	1	2	3	4	5	6	7	8	9
1	161.45	199.50	215.71	224.58	230.16	233.99	236.77	238.88	240.54
2	18.513	19.000	19.164	19.247	19.296	19.330	19.353	19.371	19.385
3	10.128	9.5521	9.2766	9.1172	9.0135	8.9406	8.8867	8.8452	8.8123
4	7.7086	9.9443	6.5914	6.3882	6.2561	6.1631	6.0942	6.0410	6.9988
5	6.6079	5.7861	5.4095	5.1922	5.0503	4.9503	4.8759	4.8183	4.7725
6	5.9874	5.1433	4.7571	4.5337	4.3874	4.2839	4.2067	4.1468	4.0990
7	5.5914	4.7374	4.3468	4.1203	3.9715	3.8660	3.7870	3.7257	3.6767
8	5.3177	4.4590	4.0662	3.8379	3.6875	3.5806	3.5005	3.4381	3.3881
9	5.1174	4.2565	3.8625	3.6331	3.4817	3.3738	3.2927	3.2296	3.1789
10	4.9646	4.1028	3.7083	3.4780	3.3258	3.2172	3.1355	3.0717	3.0204
11	4.8443	3.9823	3.5874	3.3567	3.2039	3.0946	3.0123	2.9480	2.8962
12	4.7472	3.8853	3.4903	3.2592	3.1059	2.9961	2.9134	2.8486	2.7964
13	4.6672	3.8056	3.4105	3.1791	3.0254	2.9153	2.8321	2.7669	2.7144
14	4.6001	3.7389	3.3439	3.1122	2.9582	2.8477	2.7642	2.6987	2.6458
15	4.5431	3.6823	3.2874	3.0556	2.9013	2.7905	2.7066	2.6408	2.5876
16	4.4940	3.6337	3.2389	3.0069	2.8524	2.7413	2.6572	2.5911	2.5377
17	4.4513	3.5915	3.1968	2.9647	2.8100	2.6987	2.6143	2.5480	2.4943
18	4.4139	3.5546	3.1599	2.9277	2.7729	2.6613	2.5767	2.5102	2.4563
19	4.3807	3.5219	3.1274	2.8951	2.7401	2.6283	2.5435	2.4768	2.4227
20	4.3512	3.4928	3.0984	2.8661	2.7109	2.5990	2.5140	2.4471	2.3928
21	4.3248	3.4668	3.0725	2.8401	2.6848	2.5727	2.4876	2.4205	2.3660
22	4.3009	3.4434	3.0491	2.8167	2.6613	2.5491	2.4638	2.3965	2.3419
23	4.2793	3.4221	3.0280	2.7955	2.6400	2.5277	2.4422	2.3748	2.3201
24	4.2597	3.4028	3.0088	2.7763	2.6207	2.5082	2.4226	2.3551	2.3002
25	4.2417	3.3852	2.9912	2.7587	2.6030	2.4904	2.4047	2.3371	2.2821
26	4.2252	3.3690	2.9752	2.7426	2.5868	2.4741	2.3883	2.3205	2.2655
27	4.2100	3.3541	2.9604	2.7278	2.5719	2.4591	2.3732	2.3053	2.2501
28	4.1960	3.3404	2.9467	2.7141	2.5581	2.4453	2.3593	2.2913	2.2360
29	4.1830	3.3277	2.9340	2.7014	2.5454	2.4324	2.3463	2.2783	2.2229
30	4.1709	3.3158	2.9223	2.6896	2.5336	2.4205	2.3343	2.2662	2.2107
40	4.0847	3.2317	2.8387	2.6060	2.4495	2.3359	2.2490	2.1802	2.1240
60	4.0012	3.1504	2.7581	2.5252	2.3683	2.2541	2.1665	2.0970	2.0401
120	3.9201	3.0718	2.6802	2.4472	2.2899	2.1750	2.0868	2.0164	1.9588
∞	3.8415	2.9957	2.6049	2.3719	2.2141	2.0986	2.0096	1.9384	1.8799