20

The Analysis of Variance

20.1 INTRODUCTION

Earlier, we compared two population means by using a two-sample t-test. However, we are often required to compare more than two population means simultaneously. We might be tempted to apply the two-sample t-test to all possible pairwise comparisons of means. For example, if we wish to compare 4 population means, there will be $\binom{4}{2} = 6$ separate pairs and to test the null hypothesis that all four population means are equal, would require six two-sample t-tests. Similarly, to test the null hypothesis that 10 population means are equal, we would need $\binom{10}{2} = 45$ separate two-sample t-tests. This sort of running multiple two-sample t-tests for comparing means has two disadvantages. *First*, the procedure is tedious and time consuming, and *secondly*, the overall level of significance greatly increases as the number of t-tests increases. Thus a series of two-sample t-tests is not an appropriate procedure to test the equality of several means simultaneously.

Evidently, we require a procedure for carrying out a test on several means simultaneously. One such procedure is the *analysis of variance*, introduced by Sir R.A. Fisher (1890–1962) in 1923. The *analysis of variance* (abbreviated as *ANOVA*) is a technique that partitions the total variation—a term distinct from variance and measured by the sum of squares of deviations from the mean—into its component parts, each of which is associated with a different source of variation. These component parts of variance are then analysed (hence the name analysis of variance) in such a manner that certain hypotheses can be tested. This technique is based on the facts that (i) the more the sample means differ the larger the variance becomes, and (ii) the separate components provide independent and unbiased estimates of the common population variance. The anlaysis of variance p... compares two different estimates of variance by using F-distribution to determine whether the population means are equal. The analysis of variance has been shown the most powerful and useful technique whenever the statistical data can be categorised in groups.

When each observation is classified into one sample or another according to a single criterion, we have a *one-way classification* while the classification of each observation on the basis of two criteria of classification, is called a *two-way classification*. In a similar way, a *multi-way classification* is defined. We discuss the analysis of variance procedures for the first two classifications only as a **one-way analysis of variance** and a **two-way analysis of variance** respectively.

20.2 ONE-WAY ANALYSIS OF VARIANCE

The *one-way analysis of variance* is also called the *one-variable-of-classification analysis of variance*. The data are classified into k classes or groups, etc. on the basis of a single criterion. The technical term for a class, a group or a variable of classification is *treatment*.

Suppose we have k samples of equal size r (the case of unequal sample sizes will be discussed later), selected randomly and independently, one from each of k normal populations with means $\mu_1, \mu_2, ..., \mu_k$ and common variance σ^2; and we wish to test the null hypothesis that all the k-population means are equal, *i.e.*

$$H_0 : \mu_1 = \mu_2 = ... = \mu_k$$

against the alternative hypothesis

$$H_1 : \text{Not all means are equal.}$$

Let X_{ij} denote the ith observation of the jth sample (or treatment). Then the data can be arranged as in table below:

Observation	Samples (or Treatments)						Total
	1	2	...	j	...	k	
1	X_{11}	X_{12}		X_{1j}		X_{1k}	
2	X_{21}	X_{22}		X_{2j}		X_{2k}	
.	.	.		.		.	
.	.	.		.		.	
.	.	.		.		.	
i	X_{i1}	X_{i2}		X_{ij}		X_{ik}	
.	.	.		.		.	
.	.	.		.		.	
.	.	.		.		.	
r	X_{r1}	X_{r2}		X_{rj}		X_{rk}	
Totals	$T_{.1}$	$T_{.2}$	...	$T_{.j}$	...	$T_{.k}$	$T_{..}$
Means	$X_{.1}$	$X_{.2}$	...	$X_{.j}$	...	$X_{.k}$	$X_{..}$

Here $\bar{X}_{.j}$, $\bar{X}_{..}$, $T_{.j}$ and $T_{..}$ represent the mean of the jth sample, the overall or grand mean, total of observations in the jth sample and the total of all $rk = n$ observations respectively, where the dot replaces the subscript ove which we have summed.

We test the hypothesis by comparing two independent estimates of the common population variance σ^2. The estimates of the variance can be obtained in various ways.

(i) The *first* estimate of the common population variance σ^2 is evidently obtained by pooling the k sample variances. Thus the pooled estimate of σ^2, denoted by s_p^2, is given by

$$s_p^2 = \frac{(r-1)\,s_1^2 + (r-1)\,s_2^2 + \ldots + (r-1)\,s_k^2}{n-k}, \text{ where } n = rk$$

$$= \frac{1}{n-k}\left[\sum_{i=1}^{r}(X_{i1}-\bar{X}_{.1})^2 + \sum_{i=1}^{r}(X_{i2}-\bar{X}_{.2})^2 + \ldots + \sum_{i=1}^{r}(X_{ik}-\bar{X}_{.k})^2\right]$$

$$= \frac{1}{n-k}\sum_{j=1}^{k}\sum_{i=1}^{r}(X_{ij}-\bar{X}_{.j})^2$$

Later, this estimate will be referred to as the **within samples estimate of variance** σ^2. This is an unbiased estimate regardless of the fact whether or not the null hypothesis is true.

(ii) The *second* estimate is based on the variation among the sample means assuming that all the population means are equal. Theoretically, the variance of the mean of a sample of size r is σ^2/r. Therefore, using the relation $\sigma_{\bar{x}}^2 = \frac{\sigma^2}{r}$, we get $\sigma^2 = r \cdot \sigma_{\bar{x}}^2$.

If $s_{\bar{x}}^2$ is an unbiased estimate of $\sigma_{\bar{x}}^2$, then an estimate of σ^2 will be

$$r s_{\bar{x}}^2 = r\,\frac{1}{k-1}\sum_{j=1}^{k}(\bar{X}_{.j}-\bar{X}_{..})^2$$

Let us denote this estimate by s_b^2 as later we will call it the **between sample estimate of variance** σ^2. This estimate is independent of the *within sample* estimate as it is obtained using the means of the samples. However, this estimate will become greater than the estimate obtained by pooling the sample variances when the sample means differ not true.

A *third* estimate denoted by s_T^2 can also be obtained by treating the data as one large sample consisting of n observations by the relation

$$s_T^2 = \frac{1}{n-1} \sum_{j=1}^{k} \sum_{i=1}^{r} (X_{ij} - \bar{X}_{..})^2$$

This is also an unbiased estimate of σ^2 when H_0 is true. This estimate is not of use in analysing the results but can be used to simplify the computations.

As s_p^2 and s_b^2 are the independent unbiased estimates of σ^2, so their values should not differ greatly. To detect this, *i.e.* to test the hypothesis $H_0 : \mu_1 = \mu_2 = \ldots = \mu_k$, we use the ratio

$$F = \frac{s_b^2}{s_p^2} = \frac{r \sum_{j=1}^{k} (\bar{X}_{.j} - \bar{X}_{..})^2/(k-1)}{\sum_{j=1}^{k} \sum_{i=1}^{r} (X_{ij} - \bar{X}_{.j})^2/(n-k)},$$

which, if H_0 is true, has an F-distribution with $v_1 = k-1$ and $v_2 = n-k$ degrees of freedom. We will reject H_0 when $F \geq F_\alpha(v_1, v_2)$ and conclude that the population means are *not* equal. It is important to note that the analysis of variance F-test is always a one-tailed test with rejection region located in the right tail of the F-distribution.

20.2.1 Partitioning the Sum of Squares, Equal Sample Sizes. The estimates of the common population variance σ^2 may be obtained by partitioning the total variation present in the k-samples (of equal size) taken from k normal populations. The variation of all $n = kr$ observations about the grand mean is measured by the expression

$$\sum_{j=1}^{k} \sum_{i=1}^{r} (X_{ij} - \bar{X}_{..})^2 .$$

and is called the **total sum of squares.**

To partition this total variation, let us construct the following identity:

$$X_{ij} - \bar{X}_{..} = (X_{ij} - \bar{X}_{.j}) + (\bar{X}_{.j} - \bar{X}_{..}).$$

Squaring both sides and summing over i and j, we get

$$\sum_{j=1}^{k}\sum_{i=1}^{r}(X_{ij}-\bar{X}_{..})^2 = \sum_{j=1}^{k}\sum_{i=1}^{r}(X_{ij}-\bar{X}_{.j})^2 + \sum_{j=1}^{k}\sum_{i=1}^{r}(\bar{X}_{.j}-\bar{X}_{..})^2$$

$$+\,2\sum_{j=1}^{k}\sum_{i=1}^{r}(X_{ij}-\bar{X}_{.j})(\bar{X}_{.j}-\bar{X}_{..})\,.$$

The cross-product term vanishes, because for each j

$$\sum_{i=1}^{r}(X_{ij}-\bar{X}_{.j}) = \sum_{i=1}^{r}X_{ij} - r\bar{X}_{.j} = \sum_{i=1}^{r}X_{ij} - r\left(\frac{\sum_{i=1}^{r}X_{ij}}{r}\right) = 0$$

The second term may be written as

$$\sum_{j=1}^{k}\sum_{i=1}^{r}(\bar{X}_{.j}-\bar{X}_{..})^2 = r\sum_{j=1}^{k}(\bar{X}_{.j}-\bar{X}_{..})^2$$

because the summation does not have i as a subscript and the factor not containing an i is considered as constant.

Hence we get the following sum of squares identity

$$\sum_{j=1}^{k}\sum_{i=1}^{r}(X_{ij}-\bar{X}_{..})^2 = \sum_{j=1}^{k}\sum_{i=1}^{r}(X_{ij}-\bar{X}_{.j})^2 + r\sum_{j=1}^{k}(\bar{X}_{.j}-\bar{X}_{..})^2$$

which indicates that the total variation present in the samples can be partitioned into two parts. The first part is the sum of squares of deviations of the observations from the sample mean and is called the **within (samples) sum of squares**. It is also known as the **error sum of squares**. The second part is the weighted sum of the squares of deviations of the sample means from the grand mean and is called the **between sum of squares**. We can thus represent the **sum of squares** identity symbolically by the equation

$$\text{Total } SS = \text{Within } SS + \text{Between } SS$$

It is important to note that this identity holds irrespective of the fact whether or not the null hypothesis $H_0 : \mu_1 = \mu_2 = \ldots = \mu_k$ is true.

Now, we obtain an unbiased estimate of σ^2 by dividing the **Between SS** by its degrees of freedom, *i.e.* by $k-1$ as there are k samples. A second unbiased estimate of σ^2 is obtained by dividing the **within SS** by an appropriate number of degrees of freedom which is $k(r-1)$ or $n-k$ as there are k samples, each containing r observations. These two quantities are known as **mean squares** and are denoted by MSB and MSW or MSE respectively. Therefore, to test the null hypothesis $H_0 : \mu_1 = \mu_2 = \dots = \mu_k$ against the alternative H_1 : not all means are equal, we form the ratio

$$F = \frac{MSB}{MSW} = \frac{\text{estimated variance from between SS}}{\text{estimated variance from within SS}}$$

which, if H_0 is true, has an F-distribution with $v_1 = k-1$ and $v_2 = n-k$ degrees of freedom. It has already been stated that, when H_0 is not true, MSB will be larger than σ^2, we will therefore reject H_0 at the α level of significance, if $F \geq F_\alpha(v_1, v_2)$.

20.2.2 Partitioning the Degrees of Freedom. It is interesting to note that the sum of squares identity also partitions the total number of degrees of freedom. To prove this, we take the expected values of both sides of the identity. Thus

$$E\left[\sum_{j=1}^{k}\sum_{i=1}^{r}(X_{ij}-\bar{X}_{..})^2\right] = E\left[(n-1)\frac{\sum_{j=1}^{k}\sum_{i=1}^{r}(X_{ij}-\bar{X}_{..})^2}{n-1}\right] \quad (\because n = rk)$$

$$= E\left[(n-1)\,s^2\right] = (n-1)\,\sigma^2,$$

$$E\left[\sum_{j=1}^{k}\sum_{i=1}^{r}(X_{ij}-\bar{X}_{.j})^2\right] = \sum_{j=1}^{k} E\left[(r-1)\frac{\sum_{i}(X_{ij}-\bar{X}_{.j})^2}{r-1}\right]$$

$$= \sum_{j=1}^{k} E\left[(r-1)\,s_{\omega}^2\right]$$

$$= \sum_{j=1}^{k}(r-1)\,\sigma^2 = (n-k)\,\sigma^2, \text{ and}$$

$$E\left[r\sum_{j=1}^{k}(\bar{X}_{.j}-\bar{X}_{..})^2\right] = E\left[r\sum_{j=1}^{k}\left\{(\bar{X}_{j}-\mu)-(\bar{X}_{..}-\mu)\right\}^2\right]$$

$$= E\left[r\sum_{j=1}^{k}\left\{(\bar{X}_{.j}-\mu)^2+(\bar{X}_{..}-\mu)^2-2(\bar{X}_{.j}-\mu)(\bar{X}_{..}-\mu)\right\}\right]$$

$$= E\left[\sum_{j=1}^{k} r(\bar{X}_{.j}-\mu)^2 + n(\bar{X}_{..}-\mu)^2 - 2n(\bar{X}_{..}-\mu)^2\right]$$

6

$$= \sum_{j=1}^{k} r\, E(\bar{X}_{.j} - \mu)^2 - n\, E(\bar{X}_{..} - \mu)^2$$

$$= \sum_{j=1}^{k} r \frac{\sigma^2}{r} - n \frac{\sigma^2}{n}$$

$$= k\sigma^2 - \sigma^2 = (k-1)\,\sigma^2.$$

Substituting these values in the sum of squares identity, we get

$$(n-1)\,\sigma^2 = (n-k)\,\sigma^2 + (k-1)\,\sigma^2$$

Dividing both sides by σ^2, we obtain

$$(n-1) = (n-k) + (k-1)$$

Clearly the total number of degrees of freedom is $n-1$ as there is only one restriction of computing the grand mean. The *d.f.* for k samples is $k-1$, because the mean of the sample means must equal the grand mean. Similarly, the *d.f.* for **within** SS is $n-k$, due to the k restrictions of computing the k-sample means. Hence we find that

$$\text{Total } df = \text{Within } df + \text{Between } df$$

20.2.3. The Analysis of Variance Table. The various sources of variation, degrees of freedom, the sums of squares and mean squares associated with the sources are generally shown in a table, called an **analysis of variance table** or *ANOVA* table. This table is used in testing the hypothesis that the population means differ. For one-way analysis of variance, with k samples of r observations each, the analysis of variance table is shown below:

Analysis of Variance Table

The *procedure* for testing the hypothesis $H_0 : \mu_1 = \mu_2 = \ldots = \mu_k$ using **one-way analysis of variance** (samples of equal sizes r) is given below:

(i) Formulate the null and alternative hypotheses as

$$H_0 : \mu_1 = \mu_2 = \ldots = \mu_k, \text{ and}$$

$$H_1 : \text{Not all } k \text{ means are equal.}$$

(ii) Decide upon a significance level α.

(iii) The test-statistic is

$$F = \frac{s_b^2}{s_w^2},$$

where $s_b^2 = \frac{1}{k-1}\left[r \sum_{j=1}^{k} (\bar{X}_{.j} - \bar{X}_{..})^2\right]$, and

$$s_w^2 = \frac{1}{n-k} \sum_{i=1}^{r} \sum_{j=1}^{k} (X_{ij} - \bar{X}_{.j})^2,$$

are the two estimates of the common variance σ^2, if H_0 is true. The F-statistic, if H_0 is true, has an F-distribution with $v_1 = k-1$ and $v_2 = n-k$ degrees of freedom.

(iv) Compute the necessary **Sums of Squares** and complete the analysis of variance table. Also compute F-ratio.

(v) Determine the critical region, which will consist of all values greater than or equal to $F_\alpha(k-1, n-k)$.

(vi) Decide as below:

Reject H_0 if F falls in the critical region, accept H_0 otherwise.

20.2.4. Alternative Computing Formulas. The computations of the sums of squares can be simplified as below:

$$\text{Total } SS = \sum_i \sum_j (X_{ij} - \bar{X}_{..})^2$$

$$= \sum_i \sum_j (X_{ij}^2 + \bar{X}_{..}^2 - 2 X_{ij} \bar{X}_{..})$$

$$= \sum_i \sum_j X_{ij}^2 + n\bar{X}_{..}^2 - 2\bar{X}_{..} \sum_i \sum_j X_{ij}$$

$$= \sum_i \sum_j X_{ij}^2 - n\bar{X}_{..}^2 = \sum_i \sum_j X_{ij}^2 - \frac{T_{..}^2}{n}$$

$$= \text{sum of squares of all values} - \frac{(\text{sum of all values})^2}{\text{number of values}}$$

Between $$SS = \sum_{j=1}^{k} r\,(\bar{X}_{.j} - \bar{X}_{..})^2$$

$$= r\sum_j \left(\frac{T_{.j}}{r} - \frac{T_{..}}{rk}\right)^2 = \frac{\sum_j T_{.j}^2}{r} - \frac{T_{..}^2}{n} \quad (\because\ n = rk)$$

The **Within SS** or *SSE* is usually obtained by subtracting the Between SS from **Total SS**. The term $\frac{T_{..}^2}{n}$ is generally called a *correction factor* (abbreviated as *CF*) as the deviations are taken from the grand mean. The arithmetic can further be simplified by choosing a convenient origin as all the *SS* are independent of origin.

Example 20.1 Given the data below, test the hypothesis that the means of the three populations are equal. Let $\alpha = 0.05$.

Sample 1	Sample 2	Sample 3
40	70	45
50	65	38
60	66	60
65	50	42

(i) We state our null and alternative hypotheses as

$H_0 : \mu_1 = \mu_2 = \mu_3$, *i.e.* all the three means are equal, and

H_1 : Not all three means are equal.

(ii) The significance level is set at $\alpha = 0.05$.

(iii) The test-statistic to use is

$$F = \frac{s_b^2}{s_w^2},$$

which, if H_0 is true, has an *F*-distribution with $v_1 = k - 1$ and $v_2 = n - k$ degrees of freedom.

	40 (1600) 50 (2500) 60 (3600) 65 (4225)	70 (4900) 65 (4225) 66 (4356) 50 (2500)	45 (2025) 38 (1444) 60 (3600) 42 (1764)	--- --- --- ---	8525 8169 11556 8469
$T_{\cdot j}$	215	251	185	651	36739
$T^2_{\cdot j}$	46225	63001	34225	143451	↑
$\sum_i X^2_{ij}$	11925	15981	8833	36739	← check

$$\text{Correction Factor } (C.F.) = \frac{T^2_{\cdot\cdot}}{n} = \frac{(651)^2}{12} = 35316.75$$

$$\text{Total } SS = \sum_i \sum_j X^2_{ij} - C.F.$$

$$= 36739 - 35316.75 = 1422.25$$

$$\text{Between } SS = \frac{\sum_j T^2_{\cdot j}}{r} - C.F.$$

$$= \frac{143451}{4} - 35316.75 = 546.00, \text{ and}$$

Within SS = Total SS − Between SS = 1422.25 − 546.00 = 876.25

The Analysis of Variance table is:

20.2.5. One-Way Analysis of Variance. Unequal Sample ... In the preceding sections, we discussed the one-way analysis of ... for the situation in which the k samples were all of the same ... But, generally, the sizes of the samples are not equal. Let the k ... samples be of sizes $r_1, r_2, \ldots, r_k$ respectively with $\sum_{j=1}^{k} r_j = n$. The ... of squares identity would then be written with a slight modification

$$\sum_{i=1}^{r_j}\sum_{j=1}^{k}(X_{ij}-\bar{X}_{..})^2 = \sum_{i=1}^{r_j}\sum_{j=1}^{k}(X_{ij}-\bar{X}_{.j})^2 + \sum_{j=1}^{k} r_j(\bar{X}_{.j}-\bar{X}_{..})^2$$

... formulas for computing the **Total SS** and **Between SS** are given below:

$$\text{Total} \quad SS = \sum_{j=1}^{k}\sum_{i=1}^{r_j} X_{ij}^2 - \frac{T_{..}^2}{n}$$

$$\text{Between} \quad SS = \sum_{j=1}^{k}\frac{T_{.j}^2}{r_j} - \frac{T_{..}^2}{n}.$$

The **Within SS** is obtained by subtraction as before. For degrees of freedom, we replace rk by n, therefore the respective $d.f.$ are $n-1$, $k-1$ and $n-k$. The rest of the analysis is the same.

Example 20.2. Suppose a company makes four kinds of light bulbs and it is desired to test whether there are any differences in the durabilities of the bulbs. Random samples of sizes $n_1=5$, $n_2=10$, $n_3=7$ and $n_4=5$ are selected and the following results are obtained:

$\bar{x}_1 = 14$, $\bar{x}_2 = 26$, $\bar{x}_3 = 17$, $\bar{x}_4 = 22$, $s_1^2 = 10$, $s_2^2 = 33\frac{1}{3}$, $s_3^2 = 28$, $s_4^2 = 54$,

where s^2 is the estimate of the population variance from the sample. Perform the analysis of variance to determine whether the service lives of the four kinds of bulbs do not differ from one another at $\alpha = 0.01$.

(P.U., B.A./B.Sc. 1986)

(i) The null and the alternative hypotheses corresponding to the problem that the service lives of the bulbs do not differ from one another, are formulated as

$H_0 : \mu_1 = \mu_2 = \mu_3 = \mu_4$, and

H_1 : Not all four means are equal.

(ii) The significance level is set at $\alpha = 0.01$.

(iii) The test-statistic to use is

$$F = \frac{s_b^2}{s_w^2},$$

which, if H_0 is true, has an F-distribution with $v_1 = k-1$ and $v_2 = n-k$ degrees of freedom.

(iv) Computations of sums of squares. To compute the necessary sums of squares, we have to first compute the grand (overall) mean $\bar{X}_{..}$ as

$$\bar{X}_{..} = \frac{n_1\bar{X}_1 + n_2\bar{X}_2 + n_3\bar{X}_3 + n_4\bar{X}_4}{n_1+n_2+n_3+n_4}$$

$$= \frac{5(14) + 10(26) + 7(17) + 5(22)}{5 + 10 + 7 + 5} = \frac{559}{27} = 20.7$$

Now

$$\text{Between } SS = \sum_{j=1}^{k} n_j (\bar{X}_{.j} - \bar{X}_{..})^2 \quad [\text{or } \sum n_j \bar{X}_{.j}^2 - (\sum n_j \bar{X}_j)^2/n]$$

$$= [5(14-20.7)^2 + 10(26-20.7)^2 + 7(17-20.7)^2 + 5(22-20.7)^2]$$

$$= 224.45 + 280.90 + 95.83 + 8.45 = 609.63, \text{ and}$$

$$\text{Within } SS = (n_1 - 1)s_1^2 + (n_2 - 1)s_2^2 + (n_3 - 1)s_3^2 + (n_4 - 1)s_4^2$$

$$= 4(10) + 9(33\tfrac{1}{3}) + 6(28) + 4(54)$$

$$= 40 + 300 + 168 + 216 = 724$$

The analysis of variance table becomes

Source of Variation	d.f.	Sum of Squares	Mean Square	Computed F
Between bulbs	3	609.63	203.21	6.46
Within bulbs	23	724.00	[illegible]	...

The critical region is $F \geq F_{0.01}(3, 23) = 4.76$

Conclusion. Since the computed value of $F = 6.46$ falls in the critical region, we therefore reject H_0. This implies that the data present sufficient evidence to indicate that the service lives of the four kinds of bulbs do differ from one another at 0.01 level of significance.

Example 20.3. The students in 3 classes in an elementary statistics course obtained total scores as in the table:

8-o'clock: 121, 117, 145, 108, 142, 154, 115, 81, 122, 127, 122

10-o'clock: 97, 145, 119, 139, 143, 133, 149, 107, 154.

2-o'clock: 134, 89, 108, 88, 146, 153, 130, 144, 125, 111, 87, 162.

Is there a significant difference in the scores received by the students meeting at different times of day? State completely the hypothesis you are testing and your conclusions.

(P.U., B.A./B.Sc. Hons., 1961)

(i) The *null* and *alternative hypotheses* corresponding to the problem that there is no significant difference in the scores received by students meeting at different times of day, are stated as

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

H_1 : Not all the means are equal.

(ii) Let us choose the level of significance at $\alpha = 0.05$.

(iii) The test-statistic to use is

$$F = \frac{s_b^2}{s_w^2},$$

which, if H_0 is true, has an F-distribution with $v_1 = 2$, $v_2 = 2?$ degrees of freedom.

(iv) Computations of sums of squares. To make the computational work easier, we choose our origin at $X = 100$. Then the computations are given as follows:

	Times of the Day			Total	$\sum_j X_{ij}^2$
	8-o'clock	10-o'clock	2-o'clock		
	21 (441)	-3 (9)	34 (1156)	---	1606
	17 (289)	45 (2025)	-11 (121)	---	2435
	45 (2025)	19 (361)	8 (64)	---	2450
	8 (64)	39 (1521)	-12 (144)	---	1729
	42 (1764)	43 (1849)	46 (2116)	---	5729
	54 (2916)	33 (1089)	53 (2809)	---	6814
	15 (225)	49 (2401)	30 (900)	---	3526
	-19 (361)	7 (49)	44 (1936)	---	2346
	22 (484)	54 (2916)	25 (625)	---	4025
	27 (729)	--- ---	11 (121)	---	850
	22 (484)	--- ---	-13 (169)	---	653
	-- --	--- ---	62 (3844)	---	3844
$T_{.j}$	254	286	277	817	36007
$T_{.j}^2$	64516	81796	76729	--	↑
$\sum_i X_{ij}^2$	9782	12220	14005	36007	←check

$$C.F. = \frac{T_{..}^2}{n} = \frac{(817)^2}{32} = 20859$$

Total $SS = \sum_i \sum_j X_{ij}^2 - C.F. = 36007 - 20859 = 15148$

Between $SS = \sum_j \frac{T_{.j}^2}{n_j} - C.F. = \left(\frac{64516}{11} + \frac{81796}{9} + \frac{76729}{12}\right) - C.F.$

$= (5865 + 9088 + 6394) - 20859 = 488.$

∴ Within SS = Total SS − Between SS = 14660.

The Analysis of Variance table is:

Source of Variation	$d.f.$	Sum of Squares	Mean Square	F
Between 'times'	2	488	244.0	0.48
Within 'times'	29	14,660	505.5	---
Total variation	31	15,148	---	---

(v) The critical region is $F \geq F_{0.05}(2, 29) = 3.33$.

(vi) **Conclusion.** Since the computed value of $F = 0.48$ does not fall in the critical region, so we accept H_0 and may conclude that there is no significant difference in the scores received at different times of day.