**Chapter 13**

# One-Factor Experiments - General

## 13.1 Analysis-of-Variance Technique

- In the estimation and hypothesis testing material covered in Chapters 9 and 10, we were restricted in each case to considering no more than two population parameters, for example, testing for the equality of two population means.
- In testing for the equality of two population means using independent samples from normal populations with common but unknown variance, it was necessary to obtain a pooled estimate of $\sigma^2$. This material dealing in two-sample inference represents a special case of what we call the one-factor problem.
- For example, in Exercise 10.35 on page 357, the survival time was measured for two samples of mice, where one sample received a new serum for leukemia treatment and the other sample received no treatment.
  In this case, we say that there is one factor, namely treatment, and the factor is at two levels.

- If several competing treatments were being used in the sampling process, more samples of mice would be necessary. In this case, the problem would involve one factor with more than two levels and thus more than two samples.
- In the case of more than two sample problem ($k > 2$) , it will be assumed that there are k samples from k populations.
- One very common procedure used to deal with testing population means is called the analysis of variance, or ANOVA.
- We have used the analysis-of-variance approach in regression theory to partition the total sum of squares into a portion due to regression and a portion due to error.
- **Example:**

  Suppose in an industrial experiment, an engineer is interested in how the mean absorption of moisture in concrete varies among 5 different concrete aggregates (Treatments). The samples are exposed to moisture for 48 hours. It is decided that 6 samples are to be tested for each aggregate (treatment), requiring a total of 30 samples to be tested.

  $k = 5$ (number of aggregates/treatments)

  $n = 6$ (number of observations for each aggregates/treatments)

  $k \times n = 5 \times 6 = 30$ (total number of observations)

  The data are recorded in Table 13.1.

Table 13.1: Absorption of Moisture in Concrete Aggregates

| Aggregate: (Treatment) | 1 | 2 | 3 | 4 | 5 | |
|---|---|---|---|---|---|---|
| | 551 | 595 | 639 | 417 | 563 | |
| | 457 | 580 | 615 | 449 | 631 | |
| | 450 | 508 | 511 | 517 | 522 | |
| | 731 | 583 | 573 | 438 | 613 | |
| | 499 | 633 | 648 | 415 | 656 | |
| | 632 | 517 | 677 | 555 | 679 | |
| Total | 3320 | 3416 | 3663 | 2791 | 3664 | 16,854 |
| Mean | 553.33 | 569.33 | 610.50 | 465.17 | 610.67 | 561.80 |

- The model for this situation may be set up as follows:
  There are 6 observations taken from each of 5 populations with means $\mu_1, \mu_2, \ldots, \mu_5$, respectively.

$$\mu_i = the\ mean\ absorption\ of\ moisture\ for\ the\ i-th\ aggregate\ (treament); i = 1, 2, \ldots, 5$$

We may wish to test

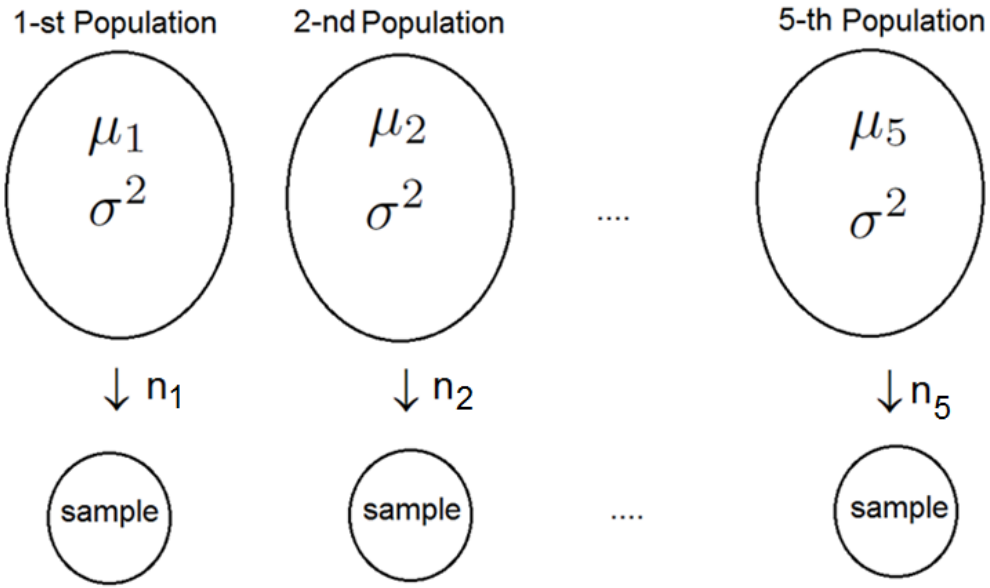
$$H_0\colon\ \mu_1 = \mu_2 = \cdots = \mu_5,$$

$$H_1\colon\ \text{At least two of the means are not equal.}$$

This is equivalent of testing:

$H_o$ : $there\ are\ no\ differences\ between\ the\ mean\ absorption\ for\ the\ 5\ aggregates\ (treaments)$

$H_1$ : $there\ are\ some\ differences\ between\ the\ mean\ absorption\ for\ the\ 5\ aggregates\ (treaments)$




We need to test:

$$H_0\colon\ \mu_1 = \mu_2 = \cdots = \mu_5,$$

$$H_1\colon\ \text{At least two of the means are not equal.}$$

We independently select a random sample of size $n_i$ form the i-th population (i=1, 2, …k).

We assume that:

- the populations are normal.
- the populations have the same variance ($\sigma^2$).

- $\sigma^2$ is unknown.
- the samples are independent.
- the sample sizes are equal ($n_1 = n_2 = \dots = n_k = n$).

Table 13.2: $k$ Random Samples

| **Treatment:** | **1** | **2** | $\cdots$ | $\boldsymbol{i}$ | $\cdots$ | $\boldsymbol{k}$ | |
|---|---|---|---|---|---|---|---|
| | $y_{11}$ | $y_{21}$ | $\cdots$ | $y_{i1}$ | $\cdots$ | $y_{k1}$ | |
| | $y_{12}$ | $y_{22}$ | $\cdots$ | $y_{i2}$ | $\cdots$ | $y_{k2}$ | |
| | $\vdots$ | $\vdots$ | | $\vdots$ | | $\vdots$ | |
| | $y_{1n}$ | $y_{2n}$ | $\cdots$ | $y_{in}$ | $\cdots$ | $y_{kn}$ | |
| Total | $Y_{1.}$ | $Y_{2.}$ | $\cdots$ | $Y_{i.}$ | $\cdots$ | $Y_{k.}$ | $Y_{..}$ |
| Mean | $\bar{y}_{1.}$ | $\bar{y}_{2.}$ | $\cdots$ | $\bar{y}_{i.}$ | $\cdots$ | $\bar{y}_{k.}$ | $\bar{y}_{..}$ |

$Y_{ij} = the\ j-th\ observation\ of\ the\ of\ the\ i-th\ treatment$

$k = number\ of\ treatments$

$n = number\ of\ observations\ for\ each\ treament$

$kn = totl\ number\ of\ observations$

$Y_{..}$ = Grand total = total of all observations = $\sum_{i=1}^{k}\sum_{j=1}^{n} y_{ij}$

$\bar{y}_{..}$ = Grand mean = mean of all observations = $\frac{Y_{..}}{kn}$

$$Y_{i.} = total\ of\ the\ observations\ of\ the\ i-th\ treatment = \sum_{j=1}^{n} y_{ij}$$

$$\bar{y}_{i.} = mean\ of\ the\ observations\ of\ the\ i-th\ treatment = \frac{Y_{i.}}{n}$$

**Two Sources of Variability in the Data**

- In the analysis-of-variance procedure, it is assumed that the variation among the aggregate averages (treatment means) is attributed to:
  - (1) variation among aggregate types (Between Groups variation / Between Treatment variation) that is, due to differences in the chemical composition of the aggregates (due to the nature of the treatments).
    (Also, it is called variation due to treatments)
  - (2) variation in absorption among observations within aggregate types (Within Group variation / Within-Sample variation / Within Treatment variation). The within aggregate variation is, of course, brought about by various causes. Perhaps humidity and temperature conditions were not kept entirely constant throughout the experiment.
    (Also, it is called variation due to error)

- we shall consider the within-sample (Within Group/ Within Treatment) variation to be chance or random variation.

- Part of the goal of the analysis of variance is to determine if the differences among the 5 sample means are what we would expect due to random variation (due to random error) alone or, rather, due to variation beyond merely random effects, i.e., differences in the chemical composition of the aggregates (differences between treatments).

## 13.3 One-Way Analysis of Variance: Completely Randomized Design (One-Way ANOVA)

Random samples of size $n$ are selected from each of $k$ populations. The $k$ different populations are classified on the basis of a single criterion such as different treatments or groups. Today the term **treatment** is used generally to refer to the various classifications, whether they be different aggregates, different analysts, different fertilizers, or different regions of the country.

### Assumptions and Hypotheses in One-Way ANOVA

It is assumed that the $k$ populations are independent and normally distributed with means $\mu_1, \mu_2, \ldots, \mu_k$ and common variance $\sigma^2$. As indicated in Section 13.2, these assumptions are made more palatable by randomization. We wish to derive appropriate methods for testing the hypothesis

$$H_0\colon\ \mu_1 = \mu_2 = \cdots = \mu_k,$$
$$H_1\colon\ \text{At least two of the means are not equal.}$$

Let $y_{ij}$ denote the $j$th observation from the $i$th treatment and arrange the data as in Table 13.2. Here, $Y_{i.}$ is the total of all observations in the sample from the $i$th treatment, $\bar{y}_{i.}$ is the mean of all observations in the sample from the $i$th treatment, $Y_{..}$ is the total of all $nk$ observations, and $\bar{y}_{..}$ is the mean of all $nk$ observations.

Table 13.2: $k$ Random Samples

| **Treatment:** | **1** | **2** | $\cdots$ | $\boldsymbol{i}$ | $\cdots$ | $\boldsymbol{k}$ | |
|---|---|---|---|---|---|---|---|
| | $y_{11}$ | $y_{21}$ | $\cdots$ | $y_{i1}$ | $\cdots$ | $y_{k1}$ | |
| | $y_{12}$ | $y_{22}$ | $\cdots$ | $y_{i2}$ | $\cdots$ | $y_{k2}$ | |
| | $\vdots$ | $\vdots$ | | $\vdots$ | | $\vdots$ | |
| | $y_{1n}$ | $y_{2n}$ | $\cdots$ | $y_{in}$ | $\cdots$ | $y_{kn}$ | |
| Total | $Y_{1.}$ | $Y_{2.}$ | $\cdots$ | $Y_{i.}$ | $\cdots$ | $Y_{k.}$ | $Y_{..}$ |
| Mean | $\bar{y}_{1.}$ | $\bar{y}_{2.}$ | $\cdots$ | $\bar{y}_{i.}$ | $\cdots$ | $\bar{y}_{k.}$ | $\bar{y}_{..}$ |

## Resolution of Total Variability into Components

Our test will be based on a comparison of two independent estimates of the common population variance $\sigma^2$. These estimates will be obtained by partitioning the total variability of our data, designated by the double summation

$$\text{SST} = \sum_{i=1}^{k}\sum_{j=1}^{n}(y_{ij} - \bar{y}_{..})^2,$$

into two components.

**Theorem 13.1:** **Sum-of-Squares Identity**

$$\sum_{i=1}^{k}\sum_{j=1}^{n}(y_{ij} - \bar{y}_{..})^2 = n\sum_{i=1}^{k}(\bar{y}_{i.} - \bar{y}_{..})^2 + \sum_{i=1}^{k}\sum_{j=1}^{n}(y_{ij} - \bar{y}_{i.})^2$$

$$\text{SST} = \text{SSA} + \text{SSE}$$

It will be convenient in what follows to identify the terms of the sum-of-squares identity by the following notation:

Three Important Measures of Variability

$$SST = \sum_{i=1}^{k}\sum_{j=1}^{n}(y_{ij} - \bar{y}_{..})^2 = \text{total sum of squares},$$

$$SSA = n\sum_{i=1}^{k}(\bar{y}_{i.} - \bar{y}_{..})^2 = \text{treatment sum of squares},$$

$$SSE = \sum_{i=1}^{k}\sum_{j=1}^{n}(y_{ij} - \bar{y}_{i.})^2 = \text{error sum of squares}.$$

The sum-of-squares identity can then be represented symbolically by the equation

$$SST = SSA + SSE.$$

The identity above expresses how between-treatment and within-treatment variation add to the total sum of squares.

Note:

Usually, we compute the error sum of squares (SSE) by difference, i. e.,

$$\boldsymbol{SSE = SST - SSA}$$

**Degrees of Freedoms:**

Degrees of freedom of SST: $df(SST) = nk - 1$

Degrees of freedom of SSA: $df(SSA) = k - 1$

Degrees of freedom of SSE: $df(SSE) = k(n-1)$

Note:

We can compute df(SSE), the degrees of freedom of the error sum of squares by difference, i. e.,

$$
\begin{aligned}
df(SSE) &= df(SST) - df(SSA) \\
&= (nk-1) - (k-1) = nk - k = k(n-1)
\end{aligned}
$$

**Mean of Squares:**

$$Mean\ of\ Square = MS = \frac{Sum\ of\ Squares}{Degrees\ of\ Freedom} = \frac{SS}{df}$$

| | |
|---|---|
| Treatment Mean Square | MSA = $s_1^2 = \dfrac{SSA}{k-1}$ |

| | |
|---|---|
| Error Mean Square | MSE = $s^2 = \dfrac{SSE}{k(n-1)}$ |

## $F$-Ratio for Testing Equality of Means

$$f = \frac{S_1^2}{S^2} = \frac{MSA}{MSE}$$

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

$H_1$: At least two of the means are not equal.

When $H_0$ is true, the ratio $f = s_1^2/s^2$ is a value of the random variable $F$ having the $F$-distribution with $k-1$ and $k(n-1)$ degrees of freedom

The null hypothesis $H_0$ is rejected at the $\alpha$-level of significance when

$$f > f_\alpha[k-1, k(n-1)].$$

The computations for an analysis-of-variance problem are usually summarized in tabular form, as shown in Table 13.3.

Note:

$$f_\alpha\big(k-1, k(n-1)\big) = f_\alpha(k-1, kn-k)$$

is the critical value or the value of f-distribution (with degrees of freedoms $\nu_1 = k-1 \; and \; \nu_2 = kn-k$ ) that leaves an area of α to the right. This value can be found from F-table.



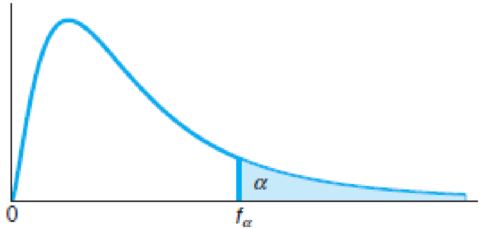



Table A.6 Critical Values of the *F*-Distribution

| | $f_{0.05}(v_1, v_2)$ | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| | $v_1$ | | | | | | | | |
| $v_2$ | 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 | 18.51 | 19.00 | 19.16 | 19.25 | 19.30 | 19.33 | 19.35 | 19.37 | 19.38 |
| 3 | 10.13 | 9.55 | 9.28 | 9.12 | 9.01 | 8.94 | 8.89 | 8.85 | 8.81 |
| 4 | 7.71 | 6.94 | 6.59 | 6.39 | 6.26 | 6.16 | 6.09 | 6.04 | 6.00 |
| 5 | 6.61 | 5.79 | 5.41 | 5.19 | 5.05 | 4.95 | 4.88 | 4.82 | 4.77 |

Table 13.3: Analysis of Variance for the One-Way ANOVA

| **Source of Variation** | **Sum of Squares** | **Degrees of Freedom** | **Mean Square** | **Computed $f$** |
|---|---|---|---|---|
| Treatments | $SSA$ | $k-1$ | $s_1^2 = \frac{SSA}{k-1}$ | $\frac{s_1^2}{s^2}$ |
| Error | $SSE$ | $k(n-1)$ | $s^2 = \frac{SSE}{k(n-1)}$ | |
| Total | $SST$ | $kn-1$ | | |

**Example 13.1:** Test the hypothesis $\mu_1 = \mu_2 = \cdots = \mu_5$ at the 0.05 level of significance for the data of Table 13.1 on absorption of moisture by various types of cement aggregates.

Table 13.1: Absorption of Moisture in Concrete Aggregates

| **Aggregate:** | **1** | **2** | **3** | **4** | **5** | |
|---|---|---|---|---|---|---|
| | 551 | 595 | 639 | 417 | 563 | |
| | 457 | 580 | 615 | 449 | 631 | |
| | 450 | 508 | 511 | 517 | 522 | |
| | 731 | 583 | 573 | 438 | 613 | |
| | 499 | 633 | 648 | 415 | 656 | |
| | 632 | 517 | 677 | 555 | 679 | |
| Total | 3320 | 3416 | 3663 | 2791 | 3664 | 16,854 |
| Mean | 553.33 | 569.33 | 610.50 | 465.17 | 610.67 | 561.80 |

Treatment = concrete aggregate

$k = 5$ (number of treatments / concrete aggregates)

$n = 6$ (number of observation for each treatment)

$\bar{y}_{..} = 561.80 \rightarrow$ grand mean = mean of all observations

$\bar{y}_{1.} = 553.33 \rightarrow$ mean of the observations of the 1-st treatment (1-st aggregate)

$\bar{y}_{2.} = 569.33 \rightarrow$ mean of the observations of the 2-nd treatment (2-nd aggregate)
$\bar{y}_{3.} = 610.50 \rightarrow$ mean of the observations of the 3-rd treatment (3-rd aggregate)
$\bar{y}_{4.} = 465.17 \rightarrow$ mean of the observations of the 4-th treatment (4-th aggregate)
$\bar{y}_{5.} = 610.67 \rightarrow$ mean of the observations of the 5-th treatment (5-th aggregate)

We need to test:

$H_o$ : $there\ are\ no\ differences\ between\ the\ mean\ absorption\ for\ the\ 5\ aggregates\ (treaments)$

$H_1$ : $there\ are\ some\ differences\ between\ the\ mean\ absorption\ for\ the\ 5\ aggregates\ (treaments)$

***Solution*:** The hypotheses are

$$H_0\colon\ \mu_1 = \mu_2 = \cdots = \mu_5,$$
$$H_1\colon\ \text{At least two of the means are not equal.}$$
$$\alpha = 0.05.$$

Critical region: $f > 2.76$ with $v_1 = 4$ and $v_2 = 25$ degrees of freedom. The sum-of-squares computations give

$$SST = 209{,}377, \quad SSA = 85{,}356,$$
$$SSE = 209{,}377 - 85{,}356 = 124{,}021.$$

These results and the remaining computations are exhibited in Figure 13.1 in the *SAS* ANOVA procedure.

Note:

$$f_\alpha\big(k-1, k(n-1)\big) = f_{0.05}(4, 25) = 2.76$$

*Table A.6 F-Distribution Probability Table*

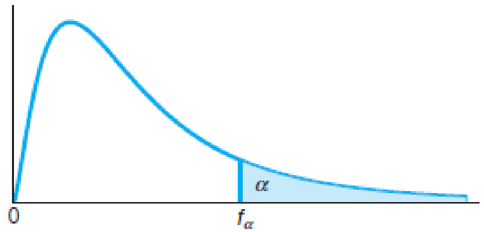



**Table A.6** Critical Values of the $F$-Distribution

| | $f_{0.05}(v_1, v_2)$ | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| | $v_1$ | | | | | | | | |
| $v_2$ | **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** | 18.51 | 19.00 | 19.16 | 19.25 | 19.30 | 19.33 | 19.35 | 19.37 | 19.38 |
| ⋮ | ⋮ | ⋮ | ⋮ | ⋮ | ⋮ | ⋮ | ⋮ | ⋮ | ⋮ |
| **24** | 4.26 | 3.40 | 3.01 | 2.78 | 2.62 | 2.51 | 2.42 | 2.36 | 2.30 |
| **25** | 4.24 | 3.39 | 2.99 | 2.76 | 2.60 | 2.49 | 2.40 | 2.34 | 2.28 |
| **26** | 4.23 | 3.37 | 2.98 | 2.74 | 2.59 | 2.47 | 2.39 | 2.32 | 2.27 |

$$\begin{aligned} SST &= \sum_{i=1}^{k}\sum_{j=1}^{n}(y_{ij} - \bar{y}_{..})^2 = \text{total sum of squares}, \\ SSA &= n\sum_{i=1}^{k}(\bar{y}_{i.} - \bar{y}_{..})^2 = \text{treatment sum of squares}, \\ SSE &= \sum_{i=1}^{k}\sum_{j=1}^{n}(y_{ij} - \bar{y}_{i.})^2 = \text{error sum of squares}. \end{aligned}$$

The sum-of-squares identity can then be represented symbolically by the equation

$$SST = SSA + SSE.$$

**(a) Calculating the total sum of squares (SST) and its degrees of freedom (df):**

$$SST = \sum_{i=1}^{k}\sum_{j=1}^{n}(y_{ij} - \bar{y}_{..})^2 = \sum_{i=1}^{5}\sum_{j=1}^{6}(y_{ij} - 561.80)^2$$

$$= (551 - 561.80)^2 + (595 - 561.80)^2 + \cdots + (679 - 561.80)^2$$

(we use all observations)

$$= 209376.8$$

df of SST:

$$df(SST) = kn - 1 = 5 * 6 - 1 = 29$$

**(b) Calculating treatment sum of squares (SSA), its degrees of freedom (df), and its mean of squares ($s_1^2$):**

$$SSA = n\sum_{i=1}^{k}(\bar{y}_{i.} - \bar{y}_{..})^2 = 6\sum_{i=1}^{5}(\bar{y}_{i.} - 561.80)^2$$

$$= 6\,[(553.33 - 561.80)^2 + (569.33 - 561.80)^2 + (610.50 - 561.80)^2 + (465.17 - 561.80)^2 + (610.67 - 561.80)^2]$$

$$= 85356.4667$$

df of SSA:

$$df(SSA) = k - 1 = 5 - 1 = 4$$

MS of treatment (MSA):

$$s_1^2 = MSA = \frac{SSA}{df(SSA)} = \frac{SSA}{k-1} = \frac{85356.4667}{4} = 21339.1167$$

**(c) Calculating error sum of squares (SSE), its degrees of freedom (df), and its mean of squares ($s^2$):**

$$SSE = By\ difference$$

$$SSE = SST - SSA = 209376.8 - 85356.4667 = 124020.3333$$

$$df(SSE) = k(n-1) = 5(6-1) = 25$$

Note: df of SSE can be calculated by difference:

df (SSE) = df(SST) – df(SSA) = 29 – 4 = 25

MS of error (MSA):

$$s^2 = MSE = \frac{SSE}{df(SSE)} = \frac{SSE}{k(n-1)} = \frac{124020.3333}{25} = 4960.8133$$

**(d) Calculating f- ratio ($\frac{s_1^2}{s^2}$):**

$$f = \frac{MSA}{MSE} = \frac{s_1^2}{s^2} = \frac{21339.1167}{4960.8133} = 4.30$$

Table 13.3: Analysis of Variance for the One-Way ANOVA

| Source of Variation | Sum of Squares | Degrees of Freedom | Mean Square | Computed $f$ |
|---|---|---|---|---|
| Treatments | $SSA$ | $k-1$ | $s_1^2 = \frac{SSA}{k-1}$ | $\frac{s_1^2}{s^2}$ |
| Error | $SSE$ | $k(n-1)$ | $s^2 = \frac{SSE}{k(n-1)}$ | |
| Total | $SST$ | $kn-1$ | | |

ANOVA table:

| Source of Variation | Sum of Squares | Degrees of Freedom | Mean Square | Computed F- ratio |
|---|---|---|---|---|
| Treatment/Aggregate (Between Treatments) | 85356.4667 | 4 | 21339.1167 | 4.30 |
| Error (Within Treatment) | 124020.3333 | 25 | 4960.8133 | |
| Total | 209376.8 | 29 | | |

For $\alpha = 0.05$, the tabulated value of F with degrees of freedoms $\upsilon_1 = 4$ and $\upsilon_2 = 25$ is:

$$f_{0.05}(4, 25) = 2.76$$

Decision:

- Since the computed F-ratio is greater than the tabulated F-value, i.e.,

$$f - ration = 4.3 > f_{0.05}(4, 25) = 2.76$$

we reject $H_o: \mu_1 = \mu_2 = \mu_3 = \mu_4 = \mu_5$

- We conclude that there are significant differences between treatments' means, which means that the concrete aggregates do not have the same mean of absorption of moisture.

**Example: (Exercise 13.1 page 518)**

Six different machines are being considered for use in manufacturing rubber seals. The machines are being compared with respect to tensile strength of the product. A random sample of four seals from each machine is used to determine whether the mean tensile strength varies from machine to machine. The following are the tensile-strength measurements in kilograms per square centimeter $\times 10^{-1}$:

| Machine | | | | | |
|---|---|---|---|---|---|
| 1 | 2 | 3 | 4 | 5 | 6 |

| | | | | | |
|---|---|---|---|---|---|
| 17.5 | 16.4 | 20.3 | 14.6 | 17.5 | 18.3 |
| 16.9 | 19.2 | 15.7 | 16.7 | 19.2 | 16.2 |
| 15.8 | 17.7 | 17.8 | 20.8 | 16.5 | 17.5 |
| 18.6 | 15.4 | 18.9 | 18.9 | 20.5 | 20.1 |

Perform the analysis of variance at the 0.05 level of significance, and indicate whether or not the mean tensile strengths differ significantly for the six machines.

**Solution:**

Let $\mu_i$ = the mean tensile strength of the i-th machine (i=1, 2, …, 6)

At the significance level $\alpha = 0.05$ , we need to test:

$$H_o: \mu_1 = \mu_2 = \mu_3 = \mu_4 = \mu_5 = \mu_6$$

against

$$H_1: at\ \ least\ \ two\ \ of\ \ the\ \ means\ are\ \ not\ \ equal$$

Machine

| | 1 | 2 | 3 | 4 | 5 | 6 | |
|---|---|---|---|---|---|---|---|
| | 17.5 | 16.4 | 20.3 | 14.6 | 17.5 | 18.3 | |
| | 16.9 | 19.2 | 15.7 | 16.7 | 19.2 | 16.2 | |
| | 15.8 | 17.7 | 17.8 | 20.8 | 16.5 | 17.5 | |
| | 18.6 | 15.4 | 18.9 | 18.9 | 20.5 | 20.1 | |
| Total = | 68.8 | 68.7 | 72.7 | 71 | 73.7 | 72.1 | Grand Total =427 |
| Mean ($\bar{y}_{i.}$) = | 17.2 | 17.175 | 18.175 | 17.75 | 18.425 | 18.025 | Grand mean ($\bar{y}_{..}$) = 17.79167 |

Treatment = Machine

$k = 6$ (number f treatments)

$n = 4$ (number of observations for each treatment)

$kn = 6 \times 4 = 24$ (total number of observations)

$\bar{y}_{..} = 17.79167$ (Grand mean = mean of all observations = sum of all observations / kn)

$\bar{y}_{i.}$ = mean of the observations of the i-th treatment (i-th machine) (i=1, 2, …, 6)
= sum of the observations of the i-th treatment (i-th machine) / n

$\bar{y}_{1.} = 17.2$ mean of the observations of the 1-st treatment (1-st machine)
$\bar{y}_{2.} = 17.175$ mean of the observations of the 2-nd treatment (2-nd machine)

$\bar{y}_{3.} = 18.175$ mean of the observations of the 3-rd treatment (3-rd machine)
$\bar{y}_{4.} = 17.75$ mean of the observations of the 4-th treatment (4-th machine)
$\bar{y}_{5.} = 18.425$ mean of the observations of the 5-th treatment (5-th machine)
$\bar{y}_{6.} = 18.025$ mean of the observations of the 5-th treatment (5-th machine)

**(a) Calculating the total sum of squares (SST) and its degrees of freedom (df):**

$$SST = \sum_{i=1}^{k}\sum_{j=1}^{n}(y_{ij} - \bar{y}_{..})^2 = \sum_{i=1}^{6}\sum_{j=1}^{4}(y_{ij} - 17.79167)^2$$

$$= (17.5 - 17.79167)^2 + (16.4 - 17.79167)^2 + \cdots$$
$$+ (20.1 - 17.79167)^2 \quad \textit{we use all sample values}$$

$$= 67.97833$$

$$df(SST) = kn - 1 = 6 * 4 - 1 = 23$$

**(b) Calculating treatment sum of squares (SSA), its degrees of freedom (df), and its mean of squares ($s_1^2$):**

$$SSA = n\sum_{i=1}^{k}(\bar{y}_{i.} - \bar{y}_{..})^2 = 4\sum_{i=1}^{5}(\bar{y}_{i.} - 17.79167)^2$$

$$= 4[(17.2 - 17.79167)^2 + (17.175 - 17.79167)^2 + (18.175 - 17.79167)^2 + (17.75 - 17.79167)^2 + (18.425 - 17.79167)^2 + (18.025 - 17.79167)^2]$$

$$= 5.338333$$

$$df(SSA) = k - 1 = 6 - 1 = 5$$

$$s_1^2 = MSA = \frac{SSA}{df(SSA)} = \frac{SSA}{k-1} = \frac{5.338333}{5} = 1.067667$$

**(c) Calculating error sum of squares (SSE), its degrees of freedom (df), and its mean of squares ($s^2$):**

$$SSE = By\ \ difference$$

$$SSE = SST - SSA = 67.97833 - 5.338333 = 62.64$$

$$df(SSE) = k(n-1) = 6(4-1) = 18$$

Note: df of SSE can be calculated by difference:

df (SSE) = df(SST) – df(SSA) = 23 – 5 = 18

$$s^2 = MSE = \frac{SSE}{df(SSE)} = \frac{SSE}{k(n-1)} = \frac{62.64}{18} = 3.48$$

**(d) Calculating F- ratio ($\frac{s_1^2}{s^2}$):**

$$f = \frac{MSA}{MSE} = \frac{s_1^2}{s^2} = \frac{1.067667}{3.48} = 0.306801$$

ANOVA table:

| Source of Variation | Sum of Squares | Degrees of Freedom | Mean Square | Computed F- ratio |
|---|---|---|---|---|
| Treatment/Machine (Between Treatments) | 5.338333 | 5 | 1.067667 | 0.306801 |
| Error (Within Treatment) | 62.64 | 18 | 3.48 | |
| Total | 67.97833 | 23 | | |

For $\alpha = 0.05$, the critical (the tabulated) value of F with degrees of freedoms $\nu_1$ =5 and $\nu_2$ =18 is:

$$f_{0.05}(5, 18) = 2.77$$

Decision:

- Since the computed F-ratio is less than the tabulated F-value, i.e,

$$f - ration = 0.306801 < \ f_{0.05}(5, 18) = \ 2.77$$

  we do not reject $H_o: \mu_1 = \mu_2 = \mu_3 = \mu_4 = \mu_5 = \mu_6$

- We conclude that there are no significant differences between treatments' means , which means that there are no significant differences between the mean tensile strength of the 6 machines.