# 9

# Testing Hypothesis for Several Population Means

SECTIONS



CHAPTER OBJECTIVES

In testing hypothesis of the population mean described in chapters 7 and 8, the number of populations was one or two. However, many cases are encountered where there are three or more population means to compare.

The analysis of variance (ANOVA) is used to test whether several population means are equal or not. The ANOVA was first published by British statistician R. A. Fisher as a test method applied to the study of agriculture, but today its principles are applied in many experimental sciences, including economics, business administration, psychology and medicine.

In section 9.1, the one-way ANOVA for single factor is introduced. In section 9.2, experimental designs for experiments are introduced. In section 9.2, the two-way ANOVA for two factors experiments is introduced.

# 9.1 Analysis of Variance for Experiments of Single Factor

- In section 8.1, we discussed how to compare means of two populations using the testing hypothesis. This chapter discusses how to compare means of several populations. There are many examples of comparing means of several populations as follows:

  - Are average hours of library usage for each grade the same?
  - Are yields of three different rice seeds equal?
  - In a chemical reaction, are response rates the same at four different temperatures?
  - Are average monthly wages of college graduates the same at three different cities?

- The group variable used to distinguish groups of the population, such as the grade or the rice, is called a **factor**.

**Definition**

**Factor**

The group variable used to distinguish groups of the population is called a **factor**.

- This section describes the one-way analysis of variance (ANOVA) which compares population means when there is a single factor. Section 9.2 describes how the experiment is designed to extract sample data. Section 9.3 describes the two-way ANOVA to compare several population means when there are two factors. Let's take a look at the following example.

**Example 9.1.1**

In order to compare the English proficiency of each grade at a university, samples were randomly selected from each grade to take the same English test, and data are as in Table 9.1.1. The right column is a calculation of the average $\overline{y}_{1\cdot}$, $\overline{y}_{2\cdot}$, $\overline{y}_{3\cdot}$, $\overline{y}_{4\cdot}$ for each grade.

Table 9.1.1 English Proficiency Score by Grade

| Grade | English Proficiency Score | | | | | | Average |
|---|---|---|---|---|---|---|---|
| 1 | 81 | 75 | 69 | 90 | 72 | 83 | $\overline{y}_{1\cdot}$ = 78.3 |
| 2 | 65 | 80 | 73 | 79 | 81 | 69 | $\overline{y}_{2\cdot}$ = 74.5 |
| 3 | 72 | 67 | 62 | 76 | 80 | | $\overline{y}_{3\cdot}$ = 71.4 |
| 4 | 89 | 94 | 79 | 88 | | | $\overline{y}_{4\cdot}$ = 87.5 |

Ex ⇨ eBook ⇨ EX090101_EnglishScoreByGrade.csv.

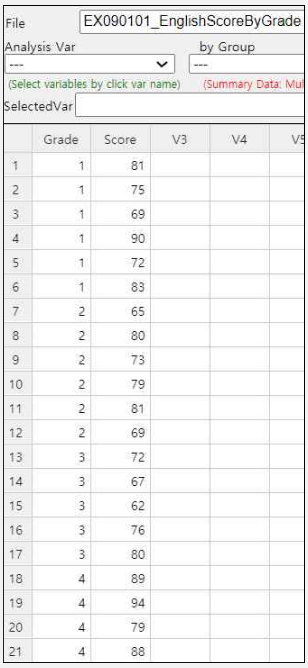
1) Using 『eStat』, draw a dot graph of test scores for each grade and compare their averages.
2) We want to test a hypothesis whether average scores of each grade are the same or not. Set up a null hypothesis and an alternative hypothesis.
3) Apply the one-way analysis of variances to test the hypothesis in question 2).
4) Use 『eStat』 to check the result of the ANOVA test.

**Example 9.1.1 Answer**

1) If you draw a dot graph of English scores by each grade, you can see whether scores of each grade are similar. If you plot the 95% confidence interval of the population mean studied in Chapter 6 on each dot graph, you can see a more detailed comparison.

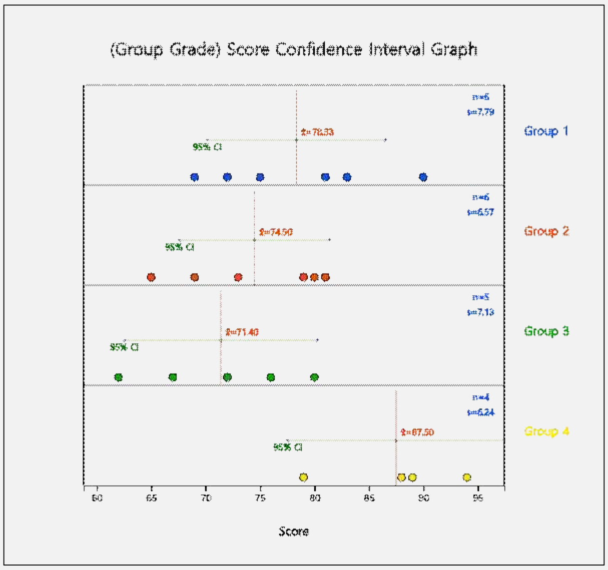
**Example 9.1.1 Answer (continued)**

- In order to draw a dot graph with data shown in Table 9.1.1 using 『eStat』, enter data on the sheet and set variable names to 'Grade' and 'Score' as shown in <Figure 9.1.1>. In the variable selection box which appears by clicking the ANOVA icon on the main menu of 『eStat』, select 'Analysis Var' as 'Score' and 'By Group' as 'Grade'. The dot graph of English scores by each grade and the 95% confidence interval are displayed as shown in <Figure 9.1.2>.




| | Grade | Score | V3 | V4 | V5 |
|---|---|---|---|---|---|
| 1 | 1 | 81 | | | |
| 2 | 1 | 75 | | | |
| 3 | 1 | 69 | | | |
| 4 | 1 | 90 | | | |
| 5 | 1 | 72 | | | |
| 6 | 1 | 83 | | | |
| 7 | 2 | 65 | | | |
| 8 | 2 | 80 | | | |
| 9 | 2 | 73 | | | |
| 10 | 2 | 79 | | | |
| 11 | 2 | 81 | | | |
| 12 | 2 | 69 | | | |
| 13 | 3 | 72 | | | |
| 14 | 3 | 67 | | | |
| 15 | 3 | 62 | | | |
| 16 | 3 | 76 | | | |
| 17 | 3 | 80 | | | |
| 18 | 4 | 89 | | | |
| 19 | 4 | 94 | | | |
| 20 | 4 | 79 | | | |
| 21 | 4 | 88 | | | |

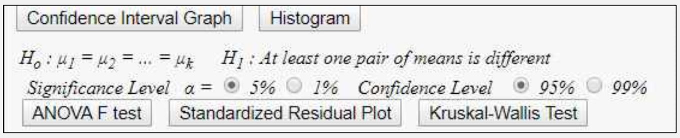
<Figure 9.1.1>
『eStat』 data input for ANOVA



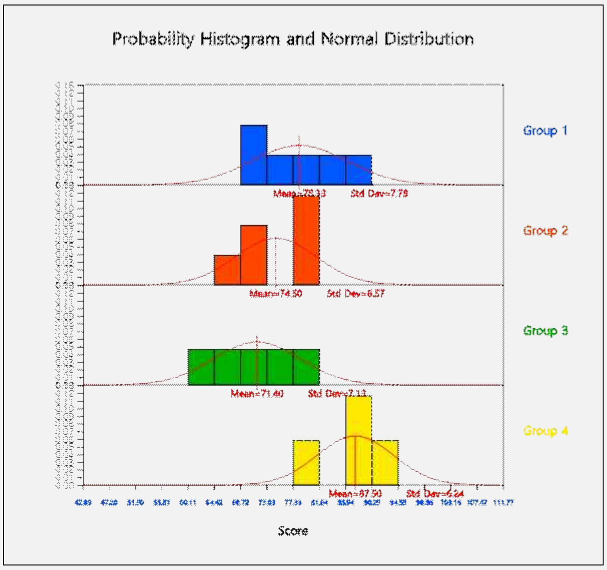

<Figure 9.1.2> 95% Confidence Interval by grade

- To review the normality of the data, pressing the [Histogram] button under this graph (<Figure 9.1.3>) will draw the histogram and normal distribution together, as shown in <Figure 9.1.4>.

**Example 9.1.1 Answer (continued)**




<Figure 9.1.3> Options of ANOVA




<Figure 9.1.4> Histogram of English score by grade

- <Figure 9.1.2> shows sample means as $\overline{y}_{1.}$= 78.3, $\overline{y}_{2.}$= 74.5, $\overline{y}_{3.}$= 71.4, $\overline{y}_{4.}$= 87.5. The sample mean of the 4$^{th}$ grader is relatively large and the order of the sample means in English is $\overline{y}_{3.}$ < $\overline{y}_{2.}$ < $\overline{y}_{1.}$ < $\overline{y}_{4.}$. $\overline{y}_{2.}$ and $\overline{y}_{3.}$ are similar, but $\overline{y}_{4.}$ is much greater than the other three. Therefore, it can be expected that the population mean $\mu_2$ and $\mu_3$ would be the same and $\mu_4$ will differ from three other population means. However, we need to test whether this difference by sample means is statistically significant.

2) In this example, the null hypothesis to test is that population means of English scores of the four grades are all the same, and the alternative hypothesis is that population means of the English scores are not the same. In other words, if $\mu_1$, $\mu_2$, $\mu_3$, $\mu_4$ are the population means of English scores for each grade, the hypothesis to test can be written as follows,

   Null hypothesis $H_0$: $\mu_1 = \mu_2 = \mu_3 = \mu_4$
   Alternative hypothesis $H_1$: at least one pair of $\mu_i$ is not the same

3) A measure that can be considered first as a basis for testing differences in multiple sample means would be the distance from each mean to the overall mean. In other words, if the overall sample mean for all 21 students is expressed as $\overline{y}_{..}$, the squared distance from each sample mean to the overall mean is as follows when the number of samples in each grade is weighted. This squared distance is called the **between sum of squares (SSB)** or the **treatment sum of squares (SSTr).**

$$\text{SSTr} = 6(78.3-\overline{y}_{..})^2 + 6(74.5-\overline{y}_{..})^2 + 5(71.4-\overline{y}_{..})^2 + 4(87.5-\overline{y}_{..})^2 = 643.633$$

   If the squared distance SSTr is close to zero, all sample means of English scores for four grades are similar.

**Example 9.1.1 Answer (continued)**

- However, this treatment sum of squares can be larger if the number of populations increases. It requires modification to become a test statistic to determine whether several population means are equal. The squared distance from each observation to its sample mean of the grade is called the **within sum of squares** (SSW) or the **error sum of squares** (SSE) as defined below.

$$\begin{aligned} \mathrm{SSE} = & (81-\overline{y}_{1\cdot})^2 + (75-\overline{y}_{1\cdot})^2 + \ \cdots \ + (83-\overline{y}_{1\cdot})^2 \\ & + (65-\overline{y}_{2\cdot})^2 + (80-\overline{y}_{2\cdot})^2 + \ \cdots \ + (69-\overline{y}_{2\cdot})^2 \\ & + (72-\overline{y}_{3\cdot})^2 + (67-\overline{y}_{3\cdot})^2 + \ \cdots \ + (80-\overline{y}_{3\cdot})^2 \\ & + (89-\overline{y}_{4\cdot})^2 + (94-\overline{y}_{4\cdot})^2 + \ \cdots \ + (88-\overline{y}_{4\cdot})^2 = 839.033 \end{aligned}$$

- If population distributions of English scores in each grade follow normal distributions and their variances are the same, the following test statistic has the $F_{3,17}$ distribution.

$$F_0 = \frac{\dfrac{\mathrm{SSTr}}{(4-1)}}{\dfrac{\mathrm{SSE}}{(21-4)}}$$

This statistic can be used to test whether population English scores of four grades are the same or not. In the test statistic, the numerator $\mathrm{SSTr}/(4-1)$ is called the **treatment mean square** (MSTr) which implies a variance between grade means. The denominator $\mathrm{SSE}/(21-4)$ is called the **error mean square** (MSE) which implies a variance within each grade. Thus, the above test statistics are based on the ratio of two variances which is why the test of multiple population means is called an analysis of variance (ANOVA).

- Calculated test statistic which is the observed F value, $F_0$, using data of English scores for each grade is as follows:

$$F_0 = \frac{\dfrac{\mathrm{SSTr}}{(4-1)}}{\dfrac{\mathrm{SSE}}{(21-4)}} = \frac{\dfrac{643.633}{3}}{\dfrac{839.033}{17}} = 4.347$$

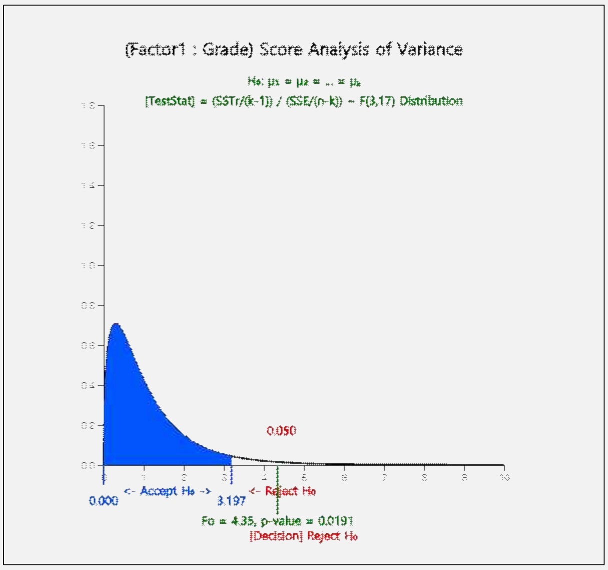
Since $F_{3,17\,;\,0.05}$ = 3.20, the null hypothesis that population means of English scores of each grade are the same, $H_0:\ \mu_1 = \mu_2 = \mu_3 = \mu_4$, is rejected at the 5% significance level. In other words, there is a difference in population means of English scores of each grade.

- The following ANOVA table provides a single view of the above calculation.

| Factor | Sum of Squares | Degree of freedom | Mean Squares | F value |
|---|---|---|---|---|
| Treatment | SSTr= 643.633 | 4-1 | MSTr = 643.633/3 | Fo = 4.347 |
| Error | SSE = 839.033 | 21-4 | MSE = 839.033/17 | |
| Total | SST =1482.666 | 20 | | |

4) In <Figure 9.1.3>, if you select the significance level of 5%, confidence level of 95%, and click [ANOVA F test] button, a graph showing the location of the test statistic in the F distribution is appeared as shown in <Figure 9.1.5>. Also, in the Log Area, the mean and confidence interval tables and test result for each grade are appeared as in <Figure 9.1.6>.

**Example 9.1.1 Answer (continued)**




<Figure 9.1.5> 『eStat』 ANOVA F test

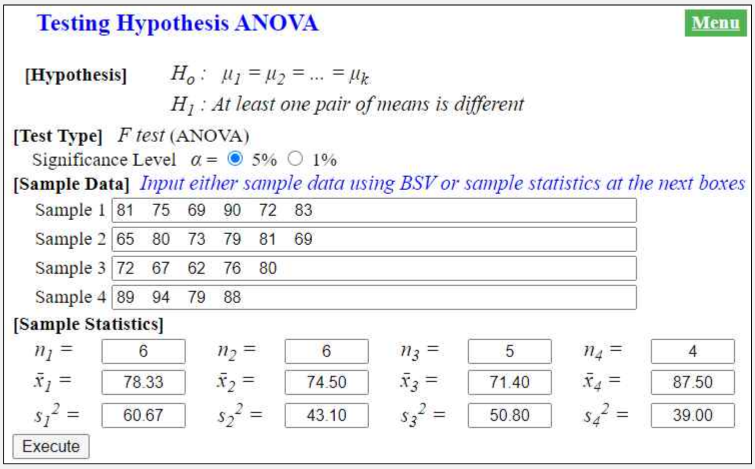
| **Statistics** | Analysis Var | Score | Group Name | Grade | | |
|---|---|---|---|---|---|---|
| Group Variable (Grade) | Observation | Mean | Std Dev | std err | Population Mean 95% Confidence Interval | Population Variance 95% Confidence Interval |
| 1 (Group 1) | 6 | 78.333 | 7.789 | 3.180 | (70.159, 86.507) | (23.638, 364.929) |
| 2 (Group 2) | 6 | 74.500 | 6.565 | 2.680 | (67.610, 81.390) | (16.793, 259.260) |
| 3 (Group 3) | 5 | 71.400 | 7.127 | 3.187 | (62.550, 80.250) | (18.235, 419.472) |
| 4 (Group 4) | 4 | 87.500 | 6.245 | 3.122 | (77.563, 97.437) | (12.516, 542.181) |
| Total | 21 | 77.333 | 8.610 | 1.879 | (73.414, 81.253) | (43.391, 154.593) |
| Missing Observations | 0 | | | | | |

| **Analysis of Variance** | | | | | |
|---|---|---|---|---|---|
| Factor | Sum of Squares | deg of freedom | Mean Squares | F value | p value |
| Treatment | 643.633 | 3 | 214.544 | 4.347 | 0.0191 |
| Error | 839.033 | 17 | 49.355 | | |
| Total | 1482.667 | 20 | | | |

<Figure 9.1.6> 『eStat』 Basic Statistics and ANOVA table

- The analysis of variance is also possible using 『eStatU』. Entering the data as in <Figure 9.1.7> and clicking the [Execute] button will have the same result as in <Figure 9.1.5>.

**Example 9.1.1 Answer (continued)**




<Figure 9.1.7> ANOVA data input at 『eStatU』

- The above example refers to two variables, the English score and grade. The variable such as the English score is called as an **analysis variable** or a **response variable.** The response variable is mostly a continuous variable. The variable used to distinguish populations such as the grade is called a group variable or a **factor** variable which is mostly a categorical variable. Each value of a factor variable Is called a **level of the factor** and the number of these levels is the number of populations to be compared. In the above example, the factor has four levels, 1$^{st}$, 2$^{nd}$, 3$^{rd}$ and 4$^{th}$ grade. The term 'response' or 'factor' is originated to analyze data through experiments in engineering, agriculture, medicine and pharmacy.
- The analysis of variance method that examines the effect of single factor on the response variable is called the one-way ANOVA. Table 9.1.2 shows the typical data structure of the one-way ANOVA when the number of levels of a factor is $k$ and the numbers of observation at each level are $n_1$, $n_2$, $\cdots$, $n_k$.

Table 9.1.2 Notation of the one-way ANOVA

| Factor | Observed values of sample | | | | Average |
|---|---|---|---|---|---|
| Level 1 | $Y_{11}$ | $Y_{12}$ | $\cdots$ | $Y_{1n_1}$ | $\overline{Y}_{1\cdot}$ |
| Level 2 | $Y_{21}$ | $Y_{22}$ | $\cdots$ | $Y_{2n_2}$ | $\overline{Y}_{2\cdot}$ |
| $\cdots$ | $\cdots$ | $\cdots$ | | $\cdots$ | $\cdots$ |
| Level $k$ | $Y_{k1}$ | $Y_{k2}$ | $\cdots$ | $Y_{kn_k}$ | $\overline{Y}_{k\cdot}$ |

- Statistical model for the one-way analysis of variance is given as follows:

$$Y_{ij} = \mu_i + \epsilon_{ij}$$

$$= \mu + \alpha_i + \epsilon_{ij}\ , \qquad i=1,2,\cdots,k\ ; \quad j=1,2,\cdots,n_i$$

$Y_{ij}$ represents the $j^{th}$ observed value of the response variable for the $i^{th}$ level of factor. The population mean of the $i^{th}$ level, $\mu_i$, is represented as $\mu+\alpha_i$ where $\mu$ is the mean of entire population and $\alpha_i$ is the **effect** of $i^{th}$ level for the response variable. $\epsilon_{ij}$ denotes an error term of the $j^{th}$ observation for the $i^{th}$ level and the all error terms are assumed independent of each other and follow the same normal distribution with the mean 0 and variance $\sigma^2$.

- The error term $\epsilon_{ij}$ is a random variable in the response variable due to reasons other than levels of the factor. For example, in the English score example, differences in English performance for each grade can be caused by other variables besides the variables of grade, such as individual study hours, gender and IQ. However, by assuming that these variations are relatively small compared to variations due to differences in grade, the error term can be interpreted as the sum of these various reasons.
- The hypothesis to test can be represented using $\alpha_i$ instead of $\mu_i$ as follows:

Null hypothesis $\quad H_0 : \alpha_1 = \alpha_2 = \cdots = \alpha_k = 0$
Alternative hypothesis $\quad H_1 :$ At least one pair of $\alpha_i$ is not equal to 0

In order to test the hypothesis, the analysis of variance table as Table 9.1.3 is used.

Table 9.1.3 Analysis of variance table of the one-way ANOVA

| Factor | Sum of Squares | Degree of freedom | Mean Squares | F value |
|---|---|---|---|---|
| Treatment<br>Error | SSTr<br>SSE | $k-1$<br>$n-k$ | MSTr = SSTr / $(k-1)$<br>MSE = SSE / $(n-k)$ | $F_0$ = MSTr/MSE |
| Total | SST | $n-1$ | | |

$(n = \sum_{i=1}^{k} n_i)$

- The three sum of squares for the analysis of variances can be described as follows. For an explanation, first define the following statistics:

$\overline{Y}_{i\cdot}$ Mean of observations at the $i^{th}$ level
$\overline{Y}_{\cdot\cdot}$ Mean of total observations

$\mathrm{SST} = \sum_{i=1}^{k}\sum_{j=1}^{n_i} (Y_{ij} - \overline{Y}_{\cdot\cdot})^2$ :
The sum of squared distances between observed values of the response variable and the mean of total observations is called the **total sum of squares** (SST).

$\mathrm{SSTr} = \sum_{i=1}^{k}\sum_{j=1}^{n_i} (\overline{Y}_{i\cdot} - \overline{Y}_{\cdot\cdot})^2$ :
The sum of squared distances between the mean of each level and the mean of total observations is called the **treatment sum of squares** (SSTr). It represents the variation between level means.

$\mathrm{SSE} = \sum_{i=1}^{k}\sum_{j=1}^{n_i} (Y_{ij} - \overline{Y}_{i\cdot})^2$ :
The sum of squared distances between observations of the $i^{th}$ level and the mean of the $i^{th}$ level which is referred to as 'within variation', and is called the error sum of squares (SSE).

- The degree of freedom of each sum of squares is determined by the following logic: The SST consists of $n$ number of squares, $(Y_{ij}-\overline{Y}_{..})^2$, but $\overline{Y}_{..}$ should be calculated first, before SST is calculated, and Hence, the degree of freedom of SST is $(n-1)$. The SSE consists of $n$ number of squares, $(Y_{ij}-\overline{Y}_{i.})^2$, but the $k$ number of values, $\overline{Y}_{1.}, \cdots, \overline{Y}_{k.}$ should be calculated first, before SSE is calculated, and Hence, the degree of freedom of SSE is $(n-k)$. The degree of freedom of SSTr is calculated as the degree of freedom of SST minus the degree of freedom of SSE which is $(k-1)$.
- In the one-way analysis of variance, the following facts are always established:

Partition of sum of squares and degrees of freedom

Sum of squares: $\mathrm{SST} = \mathrm{SSE} + \mathrm{SSTr}$

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

- The sum of squares divided by the corresponding degrees of freedom is referred to as the mean squares and Table 9.1.3 defines the **treatment mean squares (MSTr)** and **error mean squares (MSE)**. As in the meaning of the sum of squares, the treatment mean square implies the average variation between each level of the factor, and the error mean square implies the average variation within observations in each level. Therefore, if MSTr is relatively much larger than MSE, we can conclude that the population means of each level, $\mu_i$, are not the same. So by what criteria can you say it is relatively much larger?
- The calculated $F$ value, $F_0$, in the last column of the ANOVA table represents the relative size of MSTr and MSE. If the assumptions of $\epsilon_{ij}$ based on statistical theory are satisfied, and if the null hypothesis $H_0: \alpha_1 = \alpha_2 = \cdots = \alpha_k = 0$ is true, then the below test statistic follows a F distribution with degrees of freedoms $(k-1)$ and $(n-k)$.

$$F_0 = \frac{\mathrm{MSTr}}{\mathrm{MSE}} = \frac{\mathrm{SSTr}/(\mathrm{k}-1)}{\mathrm{SSE}\ /(\mathrm{n}-\mathrm{k})}$$

- Therefore, when the significance level is $\alpha$ for a test, if the calculated value $F_0$ is greater than the value of $F_{k-1,\, n-k\,;\,\alpha}$, then the null hypothesis is rejected. That is, it is determined that the population means of each factor level are not all the same.

**One-way analysis of variance $F$ test**

Null hypothesis $H_0: \alpha_1 = \alpha_2 = \cdots = \alpha_k = 0$

Alternative hypothesis $H_1:$ At least one $\alpha_i$ is not equal to 0

Test Statistic $F_0 = \frac{\mathrm{MSTr}}{\mathrm{MSE}}$

Decision Rule If $F_0 > F_{k-1,\, n-k\,;\,\alpha}$, then reject $H_0$

(Note: 『eStat』 calculates the $p$-value of this test. Hence, if the $p$-value is smaller than the significance level $\alpha$, then reject the null hypothesis. )

**[Practice 9.1.1]**

**(Plant Growth by Condition)**

Results from an experiment to compare yields (as measured by dried weight of plants) obtained under a control (leveled 'ctrl') and two different treatment conditions (leveled 'trt1' and 'trt2'). The weight data with 30 observations on control and two treatments ('crtl', 'trt1', 'trt2'), are saved at the following location of 『eStat』. Answer the following questions using『eStat』,

Ex ⇨ eBook ⇨ PR090101_Rdatasets_PlantGrowth.csv

1) Draw a dot graph of weights for each control and treatments.
2) Test a hypothesis whether the weights are the same or not. Use the 5% significance level.

## 9.1.1 multiple comparisons

- If the F test of the one-way ANOVA does not show a significant difference between each level of the factor, it can be concluded that there is no difference between each level of populations. However, if you conclude that there are significant differences between each level as shown in [Example 9.1.1], you need to examine which levels are different from each other.
- The analysis of differences between population means after ANOVA requires several tests for the mean difference to be performed simultaneously and it is called as the **multiple comparisons**. The hypothesis for the multiple comparisons to test whether the level means, $\mu_i$ and $\mu_j$, are equal is as follows:

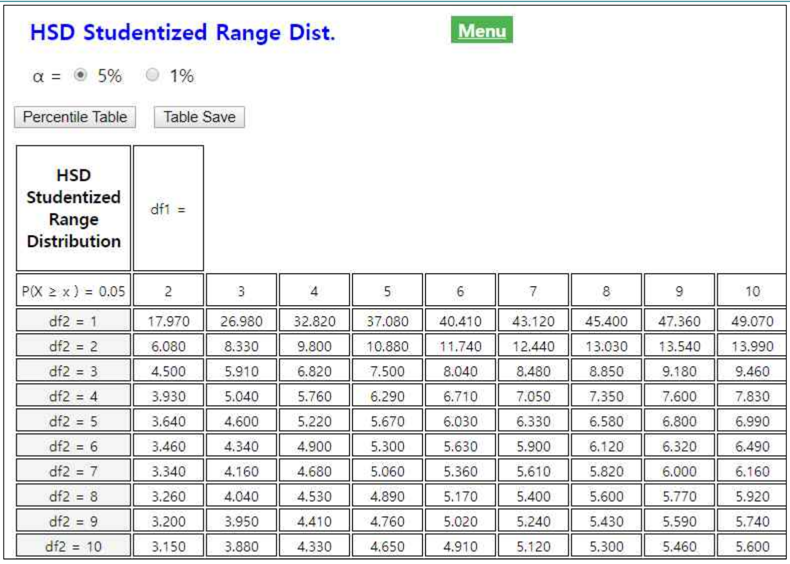
$$H_0 : \mu_i = \mu_j,\ H_1 : \mu_i \neq \mu_j \qquad i = 1,2,...,k-1;\ j = i+1, i+2, \ldots, k$$

It means that there are ${}_kC_2$ tests to be done simultaneously for the multiple comparisons if there are $k$ levels of the factor.
- There are many multiple comparisons tests, but Tukey's Honestly Significant Difference (HSD) test is most commonly used. The statistic for Tukey's HSD test to compare means $\mu_i$ and $\mu_j$ is the sample mean difference $\overline{y}_{i\cdot} - \overline{y}_{j\cdot}$ and the decision rule to test $H_0 : \mu_i = \mu_j$ is as follows:

If $|\overline{y}_i - \overline{y}_j| > HSD_{ij}$, then reject $H_0$

where $\mathrm{HSD_{ij}} = \mathrm{q_{k,n-k;\alpha}} \cdot \sqrt{\frac{1}{2}\left(\frac{1}{\mathrm{n_i}} + \frac{1}{\mathrm{n_j}}\right)\mathrm{MSE}}$,

$n_i$ and $n_j$ are the number of samples (repetitions) in $i^{th}$ level and $j^{th}$ level, $\mathrm{MSE}$ is the mean squared error, $q_{k,\ n-k\,;\,\alpha}$ is the right tail $100 \times \alpha$ percentile of the studentized range distribution with parameter $k$ and $n-k$ degrees of freedom. (It can be found at『eStatU』(<Figure 9.1.8>)).




| HSD Studentized Range Distribution | df1 = | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| P(X ≥ x ) = 0.05 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 |
| df2 = 1 | 17.970 | 26.980 | 32.820 | 37.080 | 40.410 | 43.120 | 45.400 | 47.360 | 49.070 |
| df2 = 2 | 6.080 | 8.330 | 9.800 | 10.880 | 11.740 | 12.440 | 13.030 | 13.540 | 13.990 |
| df2 = 3 | 4.500 | 5.910 | 6.820 | 7.500 | 8.040 | 8.480 | 8.850 | 9.180 | 9.460 |
| df2 = 4 | 3.930 | 5.040 | 5.760 | 6.290 | 6.710 | 7.050 | 7.350 | 7.600 | 7.830 |
| df2 = 5 | 3.640 | 4.600 | 5.220 | 5.670 | 6.030 | 6.330 | 6.580 | 6.800 | 6.990 |
| df2 = 6 | 3.460 | 4.340 | 4.900 | 5.300 | 5.630 | 5.900 | 6.120 | 6.320 | 6.490 |
| df2 = 7 | 3.340 | 4.160 | 4.680 | 5.060 | 5.360 | 5.610 | 5.820 | 6.000 | 6.160 |
| df2 = 8 | 3.260 | 4.040 | 4.530 | 4.890 | 5.170 | 5.400 | 5.600 | 5.770 | 5.920 |
| df2 = 9 | 3.200 | 3.950 | 4.410 | 4.760 | 5.020 | 5.240 | 5.430 | 5.590 | 5.740 |
| df2 = 10 | 3.150 | 3.880 | 4.330 | 4.650 | 4.910 | 5.120 | 5.300 | 5.460 | 5.600 |

<Figure 9.1.8>『eStatU』HSD percentile table

**Example 9.1.2**

In [Example 9.1.1], the analysis variance of English scores by the grade concluded that the null hypothesis was rejected and the average English scores for each grade were not all the same. Now let's apply the multiple comparisons to check where the differences exist among each school grade with the significance level of 5%. Use 『eStat』 to check the result.

**Answer**

- The hypothesis of the multiple comparisons is $H_0 :\ \mu_i = \mu_j$, $\ H_1 :\ \mu_i \neq \mu_j$ and the decision rule is as follows:

  ‘If $|\overline{y}_i - \overline{y}_j| \ > \ HSD_{ij}$, then reject $H_0$.’

  Since there are four school grades ($k=4$), ${}_4C_2$ = 6 multiple comparisons are possible as follows. The 5 percentile from the right tail of HSD distribution which is used to test is $q_{k, n-k\,;\,\alpha} = q_{4, 21-4\,;\,0.05} = 4.02$.

  1) $H_0 :\ \mu_1 = \mu_2 \quad H_1 :\ \mu_1 \neq \mu_2$

  $|\overline{y}_1 - \overline{y}_2| = |78.3 - 74.5| = 3.8$

  $$\mathrm{HSD}_{12} \ = \ \mathrm{q}_{\mathrm{k,\,n-k\,;\,0.05}} \cdot \sqrt{\frac{1}{2}(\frac{1}{\mathrm{n}_1} + \frac{1}{\mathrm{n}_2})\mathrm{MSE}}$$

  $$= \ q_{4,\,21-4\,;\,0.05} \cdot \sqrt{\frac{1}{2}(\frac{1}{6} + \frac{1}{6})49.355} \ = \ 11.530$$

  Therefore, accept $H_0$.

  2) $H_0 :\ \mu_1 = \mu_3 \quad H_1 :\ \mu_1 \neq \mu_3$

  $|\overline{y}_1 - \overline{y}_3| = |78.3 - 71.4| = 6.9$

  $$\mathrm{HSD}_{13} \ = \ \mathrm{q}_{\mathrm{k,\,n-k\,;\,0.05}} \cdot \sqrt{\frac{1}{2}(\frac{1}{\mathrm{n}_1} + \frac{1}{\mathrm{n}_3})\mathrm{MSE}}$$

  $$= \ q_{4,\,21-4\,;\,0.05} \cdot \sqrt{\frac{1}{2}(\frac{1}{6} + \frac{1}{5})49.355} \ = \ 12.092$$

  Therefore, accept $H_0$.

**Example 9.1.2 Answer (continued)**

3) $H_0:\ \mu_1 = \mu_4 \quad H_1:\ \mu_1 \neq \mu_4$

$|\overline{y}_1 - \overline{y}_4| = |78.3 - 88.5| = 10.2$

$$\mathrm{HSD}_{14} = \mathrm{q}_{\mathrm{k,n-k\,;\,0.05}} \cdot \sqrt{\frac{1}{2}(\frac{1}{\mathrm{n}_1} + \frac{1}{\mathrm{n}_4})\mathrm{MSE}}$$

$$= q_{4,21-4\,;\,0.05} \cdot \sqrt{\frac{1}{2}(\frac{1}{6} + \frac{1}{4})49.355} = 12.891$$

Therefore, accept $H_0$.

4) $H_0:\ \mu_2 = \mu_3 \quad H_1:\ \mu_2 \neq \mu_3$

$|\overline{y}_2 - \overline{y}_3| = |74.5 - 71.4| = 3.1$

$$\mathrm{HSD}_{23} = \mathrm{q}_{\mathrm{k,n-k\,;\,0.05}} \cdot \sqrt{\frac{1}{2}(\frac{1}{\mathrm{n}_2} + \frac{1}{\mathrm{n}_3})\mathrm{MSE}}$$

$$= q_{4,21-4\,;\,0.05} \cdot \sqrt{\frac{1}{2}(\frac{1}{6} + \frac{1}{5})49.355} = 12.092$$

Therefore, accept $H_0$.

5) $H_0:\ \mu_2 = \mu_4 \quad H_1:\ \mu_2 \neq \mu_4$

$|\overline{y}_2 - \overline{y}_4| = |74.5 - 88.5| = 14$

$$\mathrm{HSD}_{24} = \mathrm{q}_{\mathrm{k,n-k\,;\,0.05}} \cdot \sqrt{\frac{1}{2}(\frac{1}{\mathrm{n}_2} + \frac{1}{\mathrm{n}_4})\mathrm{MSE}}$$

$$= q_{4,21-4\,;\,0.05} \cdot \sqrt{\frac{1}{2}(\frac{1}{6} + \frac{1}{4})49.355} = 12.891$$

Therefore, reject $H_0$.

6) $H_0:\ \mu_3 = \mu_4 \quad H_1:\ \mu_3 \neq \mu_4$

$|\overline{y}_3 - \overline{y}_4| = |71.4 - 88.5| = 17.1$

$$\mathrm{HSD}_{34} = \mathrm{q}_{\mathrm{k,n-k\,;\,0.05}} \cdot \sqrt{\frac{1}{2}(\frac{1}{\mathrm{n}_3} + \frac{1}{\mathrm{n}_4})\mathrm{MSE}}$$

$$= q_{4,21-4\,;\,0.05} \cdot \sqrt{\frac{1}{2}(\frac{1}{5} + \frac{1}{4})49.355} = 13.396$$

Therefore, reject $H_0$.

- The result of the above multiple comparisons shows that there is a difference between $\mu_3$ and $\mu_4$, $\mu_2$and $\mu_4$ as can be seen in the dot graph with average in <Figure 9.1.1>. It also shows that $\mu_1$ has no significant difference from other means.
- If you click [Multiple Comparison] in the options of the ANOVA as in <Figure 9.1.3>, 『eStat』 shows the result of Tukey's multiple comparisons as shown in <Figure 9.1.9>. 『eStat』 also shows the mean difference and 95% HSD value for the sample mean combination after rearranging levels of rows and columns in ascending order of the sample means.
- The next table shows that, if the HSD test result for the combination of the two levels is significant with the 5% significance level, then * will be marked and if it is significant with the 1% significance level, then ** will be marked, if it is not significant, then the cell is left blank.

**Example 9.1.2 Answer (continued)**

| Multiple Comparison | Analysis Var | (Score) | Group Name | (Grade) |
|---|---|---|---|---|
| Mean Difference (95%HSD) | 1 (Group 1) 78.33 | 2 (Group 2) 74.50 | 3 (Group 3) 71.40 | 4 (Group 4) 87.50 |
| 1 (Group 1) 78.33 | | 3.83 (11.53) | 6.93 (12.09) | 9.17 (12.89) |
| 2 (Group 2) 74.50 | 3.83 (11.53) | | 3.10 (12.09) | 13.00 (12.89) |
| 3 (Group 3) 71.40 | 6.93 (12.09) | 3.10 (12.09) | | 16.10 (13.40) |
| 4 (Group 4) 87.50 | 9.17 (12.89) | 13.00 (12.89) | 16.10 (13.40) | |
| | | | | |
| Testing Means * 95%, ** 99% | 3 (Group 3) 71.40 | 2 (Group 2) 74.50 | 1 (Group 1) 78.33 | 4 (Group 4) 87.50 |
| 3 (Group 3) 71.40 | | | | * |
| 2 (Group 2) 74.50 | | | | * |
| 1 (Group 1) 78.33 | | | | |
| 4 (Group 4) 87.50 | * | * | | |

<Figure 9.1.9> HSD Multiple Comarisons

- For the analysis of mean differences, confidence intervals for each level may also be used. <Figure 9.1.2> shows the 95% confidence interval for the mean for each level. This confidence interval is created using the formula described in Chapter 6, but the only difference is that the estimate of the variance for the error, $\sigma^2$, is the pooled variance using overall observations rather than the sample variance of observed values at each level. In the ANOVA table, MSE is the pooled variance.
- In post-analysis using these confidence intervals, there is a difference between means if the confidence intervals are not overlapped, so the same conclusion can be obtained as in the previous HSD test.

**[Practice 9.1.2]**

By using the data of [Practice 9.1.1]

Ex ⇨ eBook ⇨ PR090101_Rdatasets_PlantGrowth.csv

apply the multiple comparisons to check where differences exist among Control and two treatments with the significance level of 5%. Use『eStat』.

## 9.1.2 Residual Analysis

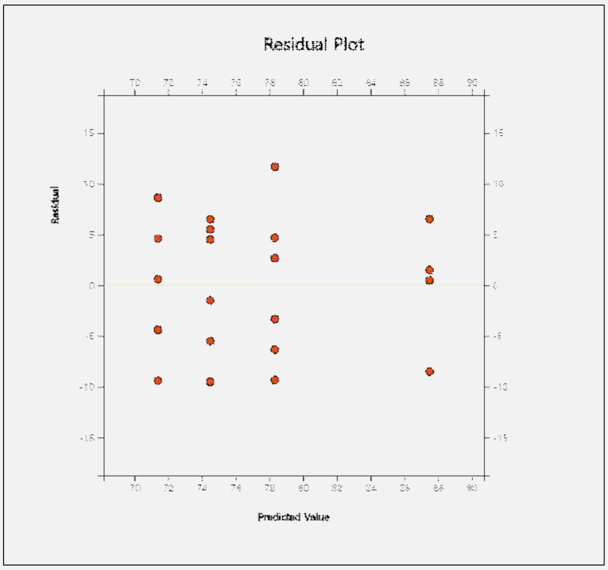
- Another statistical analysis related to the ANOVA is a **residual analysis**. Various hypothesis tests in the ANOVA are performed on the condition that assumptions hold about the error term $\epsilon_{ij}$. Assumptions about error terms include independence ($\epsilon_{ij}$ are independent of each other), homoscedasticity (each variance of $\epsilon_{ij}$ is constant as $\sigma^2$), normality (each $\epsilon_{ij}$ is normally distributed), etc. The validity of these assumptions should always be investigated. However, since $\epsilon_{ij}$ can not be observed, the residual as the estimate of $\epsilon_{ij}$ is used to check the assumptions. The residuals in the ANOVA are defined as the deviations used in the equation of the error sum of squares, for example, $(Y_{ij} - \overline{Y}_{i.})$ in the one-way analysis of variance.

**Example 9.1.3**

In [Example 9.1.1] of English score comparison by the grade, apply the residual analysis using『eStat』.

**Answer**

- If you click on [Standardized Residual Plot] of the ANOVA option in <Figure 9.1.3>, a scatter plot of residuals versus fitted values appears as shown in <Figure 9.1.10>. In this scatter plot, if the residuals show no unusual tendency around zero and appear randomly, then the assumptions of independence and homoscedasticity are valid. There is no unusual tendency in this scatter plot. Normality of the residuals can be checked by drawing the histogram of residuals.




<Figure 9.1.10> Residual plot of the ANOVA

**[Practice 9.1.3]**

By using the data of [Practice 9.1.1]

Ex ⇨ eBook ⇨ PR090101_Rdatasets_PlantGrowth.csv

apply the residual analysis using『eStat』.

## 9.2 Design of Experiments for Sampling

- Data such as English scores by the grade in [Example 9.1.1] are not so difficult to collect samples from each of the grade population. However, obtaining samples through experiments such as engineering, medicine, or agriculture are often difficult to collect a large number of samples due to the influence of many other external factors, and should be very cautious about sampling. This section discusses how to design experiments for collecting small number of data from experiments.

### 9.2.1 Completely Randomized Design

- In order to identify the differences accurately that may exist among each level of a factor, you should design experiments such as little influence from other factors. One method to do this is to make the whole experiments random. For example, consider experiments to compare a fuel mileage per one liter of gasoline for three types of cars A, B and C. We want to measure the fuel mileage for five different cars of each type. One driver may try to drive all 15 cars. However, if only five cars can be measured per day, the measurement will take place over a total of three days. In this case, changes in daily weather, wind speed and wind direction can influence the fuel mileage which makes it a question of which car should be measured for fuel mileage on each day.
- If five drivers (1, 2, 3, 4, 5) plan to drive the car to measure the fuel mileage of all cars a day, the fuel mileage of the car may be affected by the driver. One solution would be to allocate 15 cars randomly to five drivers and then to randomize the sequence of experiments as well. For example, each car is numbered from 1 to 15 and then, the experiment of the fuel mileage is conducted in the order of numbers that come out using drawing a random number. Such an experiment would reduce the likelihood of differences caused by external factors such as the driver, daily wind speed and wind direction, because randomized experiments make all external factors equally affecting the all observed measurement values. This method of experiments is called a **completely randomized design of experiments**. Table 9.2.1 shows an example allocation of experiments by this method. Symbols A, B and C represent the three types of cars.

Table 9.2.1 Example of completely randomized design of experiments

| Driver | 1 | 2 | 3 | 4 | 5 |
|---|---|---|---|---|---|
| Car Type | B | A | B | C | A |
| | B | C | A | A | C |
| | C | B | A | B | C |

- In general, in order to achieve the purpose of the analysis of variance, it is necessary to plan experiments thoroughly in advance for obtaining data properly. The completely randomized design method explained as above is studied in detail at the Design of Experiments area in Statistics. From the standpoint of the experimental design, the one-way analysis of variance technique is called an analysis of the single factor design.

### 9.2.2 Randomized Block Design

- In the experiments of completely randomized design for measuring the fuel mileage explained in the previous section, 15 cars were randomly allocated to five drivers. However, one example allocation as inTable 9.2.1 shows a problem of this completely randomized design. For example, Driver 1 will only experiment with B and C types of cars and Driver 3 will only experiment A and B types of cars so that the variable between drivers will not be averaged in the test. Thus, if there is a significant variation between drivers for measuring the fuel mileage, the error term of the analysis of variance may not be a simple experimental error. In order to eliminate this problem, each driver may be required to experiment with each type of the car at least once which is known as a **randomized block design**. Table 9.2.2 shows an example of possible allocation in this case. In this table, the values in parentheses are the values of the observed fuel mileage.

Table 9.2.2 Example of randomized block design

| Driver | 1 | 2 | 3 | 4 | 5 |
|---|---|---|---|---|---|
| Car Type (gas mileage) | A(22.4)<br>C(20.2)<br>B(16.3) | B(12.6)<br>C(15.2)<br>A(16.1) | C(18.7)<br>A(19.7)<br>B(15.9) | A(21.1)<br>B(17.8)<br>C(18.9) | A(24.5)<br>C(23.8)<br>B(21.0) |

- Table 9.2.2 shows that the total observed values are divided into five groups by driver, called blocks so that they have the same characteristics. The variable representing blocks, such as the driver, is referred to as a **block variable**. A block variable is considered generally if experimental results are influenced significantly by this variable which is different from the factor. For example, when examining the yield resulting from rice variety, if the fields of the rice paddy used in the experiment do not have the same fertility, divide the fields into several blocks which have the same fertility and then all varieties of rice are planted in each block of the rice paddy. This would eliminate the influence of the rice paddy which have different fertility and would allow for a more accurate examination of the differences in yield between rice varieties.
- Statistical model of the randomized block design with $b$ blocks can be represented as follows:

$$Y_{ij} = \mu + \alpha_i + B_j + \epsilon_{ij} \quad i = 1,2,\cdots,k \quad j = 1,2,\cdots,b$$

In this equation, $B_j$ is the effect of $j^{th}$ level of the block variable to the response variable. In the randomized block design, the variation resulting from the difference between levels of the block variable can be separated from the error term of the variation of the factor independently. In the randomized block design, the total variation is divided into as follows:

$$Y_{ij} - \overline{Y}_{..} = (Y_{ij} - \overline{Y}_{i.} - \overline{Y}_{.j} + \overline{Y}_{..}) + (\overline{Y}_{i.} - \overline{Y}_{..}) + (\overline{Y}_{.j} - \overline{Y}_{..})$$

- If you square both sides of the equation above and then combine for all $i,j$, you can obtain several sums of squares as in the one-way analysis of variance as follows:

Total sum of squares, degrees of freedom $bk-1$

$$\mathrm{SST} = \sum_{i=1}^{k}\sum_{j=1}^{b}(Y_{ij} - \overline{Y}_{..})^2$$

Error sum of squares, degrees of freedom $(b-1)(k-1)$

$$\mathrm{SSE} = \sum_{i=1}^{k}\sum_{j=1}^{b}(Y_{ij} - \overline{Y}_{i.} - \overline{Y}_{.j} + \overline{Y}_{..})^2$$

Treatment sum of squares, degrees of freedom $k-1$

$$\mathrm{SSTr} = \sum_{i=1}^{k}\sum_{j=1}^{b}(\overline{Y}_{i.} - \overline{Y}_{..})^2 \quad = \quad b\sum_{i=1}^{k}(\overline{Y}_{i.} - \overline{Y}_{..})^2$$

Block sum of squares, degrees of freedom $b-1$

$$\mathrm{SSB} = \sum_{i=1}^{k}\sum_{j=1}^{b}(\overline{Y}_{.j} - \overline{Y}_{..})^2 \quad = \ k\sum_{j=1}^{b}(\overline{Y}_{.j} - \overline{Y}_{..})^2$$

- The following facts are always established in the randomized block design.

> Division of the sum of squares and degrees of freedom
> Sum of squares : $\mathrm{SST} = \mathrm{SSE} + \mathrm{SSTr} + \mathrm{SSB}$
> Degrees of freedom : $bk-1 = (b-1)(k-1) + (k-1) + (b-1)$

- Table 9.2.3 shows the ANOVA table of the randomized block design. In this ANOVA table, if you combine the sum of squares and degrees of freedom of the block variable and the error variation, it becomes the sum of squares and degrees of freedom of the error term in the one-way ANOVA table 9.1.3.

Table 9.2.3 Analysis of Variance Table of the randomized block design

| Variation | Sum of Squares | Degrees of freedom | Mean Squares | F value |
|---|---|---|---|---|
| Treatment | $\mathrm{SSTr}$ | $k-1$ | $\mathrm{MSTr} = \dfrac{\mathrm{SSTr}}{\mathrm{k}-1}$ | $F_0 = \dfrac{MSTr}{MSE}$ |
| Block | $\mathrm{SSB}$ | $b-1$ | $\mathrm{MSB} = \dfrac{\mathrm{SSB}}{\mathrm{b}-1}$ | |
| Error | $\mathrm{SSE}$ | $(b-1)(k-1)$ | $\mathrm{MSE} = \dfrac{\mathrm{SSE}}{(\mathrm{b}-1)(\mathrm{k}-1)}$ | |
| Total | $\mathrm{SST}$ | $bk-1$ | | |

- In the randomized block design, the entire experiments are not randomized unlike the completely randomized design, but only the experiments in each block are randomized.
- Another important thing to note in the randomized block design is that, although the variation of the block variable was separated from the error variation, the main objective is to test the difference between levels of a factor as in the one-way analysis of variance. The test for differences between the levels of the block variable is not important, because the block variable is used to reduce the error variation and to make the test for differences between the levels of the factor more accurate.
- In addition, the error mean square (MSE) does not always decrease, because although the block variation is separated from the error variation of the one-way analysis of variance, the degrees of freedom are also reduced.

**Example 9.2.1**

Table 9.2.4 is the rearrangement of the fuel mileage data in Table 9.2.2 measured by five drivers and car types.

Table 9.2.4 Fuel mileage data by five drivers and three car types

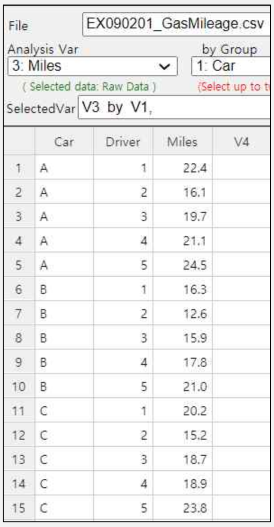
| | Drive | 1 | 2 | 3 | 4 | 5 | Average($\overline{y}_{i\cdot}$) |
|---|---|---|---|---|---|---|---|
| Car Type | A | 22.4 | 16.1 | 19.7 | 21.1 | 24.5 | 20.76 |
| | B | 16.3 | 12.6 | 15.9 | 17.8 | 21.0 | 16.72 |
| | C | 20.2 | 15.2 | 18.7 | 18.9 | 23.8 | 19.36 |
| Average($\overline{y}_{\cdot j}$) | | 19.63 | 14.63 | 18.10 | 19.27 | 23.10 | 18.947 |

Ex ⇨ eBook ⇨ EX090201_GasMilage.csv

1) Assuming that this data have been measured by the completely randomized design, use 『eStat』 to do the analysis of variance whether the three car types have the same fuel mileage.
2) Assuming that this data have been measured by the randomized block design, use 『eStat』 to do the analysis of variance whether the three car types have the same fuel mileage.

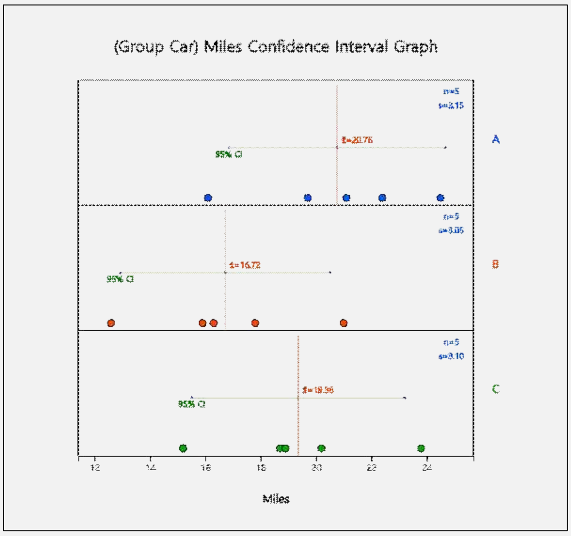
**Answer**

1) In 『eStat』, enter data as shown in <Figure 9.2.1> and click the icon of analysis of variance. Select 'Analysis Var' as Miles and 'By Group' as Car in the variable selection box, then the confidence interval graph for each type of cars will appear such as <Figure 9.2.2>.


File: EX090201_GasMileage.csv
Analysis Var: 3: Miles; by Group: 1: Car
( Selected data: Raw Data ) (Select up to t
SelectedVar: V3 by V1,

| | Car | Driver | Miles | V4 |
|---|---|---|---|---|
| 1 | A | 1 | 22.4 | |
| 2 | A | 2 | 16.1 | |
| 3 | A | 3 | 19.7 | |
| 4 | A | 4 | 21.1 | |
| 5 | A | 5 | 24.5 | |
| 6 | B | 1 | 16.3 | |
| 7 | B | 2 | 12.6 | |
| 8 | B | 3 | 15.9 | |
| 9 | B | 4 | 17.8 | |
| 10 | B | 5 | 21.0 | |
| 11 | C | 1 | 20.2 | |
| 12 | C | 2 | 15.2 | |
| 13 | C | 3 | 18.7 | |
| 14 | C | 4 | 18.9 | |
| 15 | C | 5 | 23.8 | |

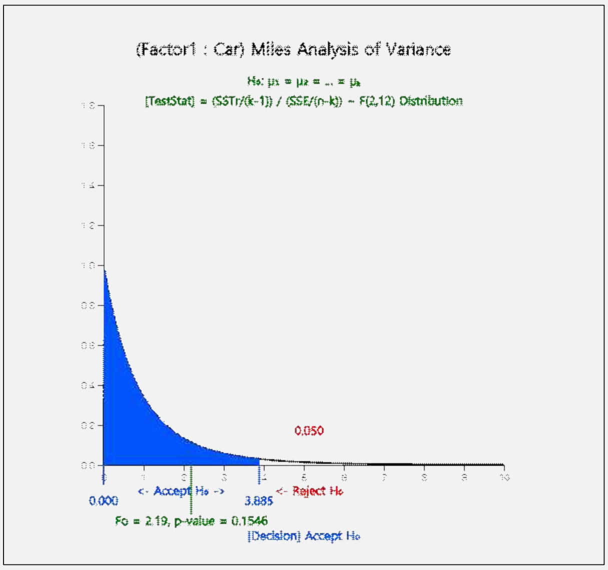
<Figure 9.2.1> Data input for randomized block design for 『eStat』 ANOVA




<Figure 9.2.2> Dot graph and 95% confidence interval for population mean of each car type

- Click the [ANOVA F-test] button in the option below the graph to reveal the ANOVA graph as in <Figure 9.2.3> and the ANOVA table as in <Figure 9.2.4>. The result of the ANOVA is that there is no difference in fuel mileage between the cars of each company. The same is true for the multiple comparisons tests in <Figure 9.2.5>.

**Example 9.2.1 Answer (continued)**




<Figure 9.2.3> ANOVA of gas milage

| Analysis of Variance | | | | | |
|---|---|---|---|---|---|
| Factor | Sum of Squares | deg of freedom | Mean Squares | F value | p value |
| Treatment | 42.085 | 2 | 21.043 | 2.190 | 0.1546 |
| Error | 115.312 | 12 | 9.609 | | |
| Total | 157.397 | 14 | | | |

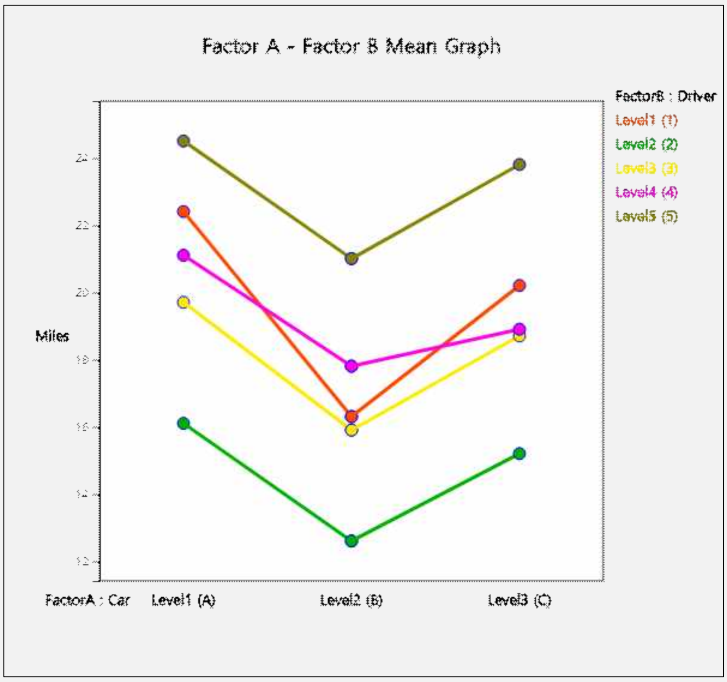
<Figure 9.2.4> ANOVA table of gas milage

| Multiple Comparison | Analysis Var | (Miles) | Group Name | (Car) |
|---|---|---|---|---|
| Mean Difference (95%HSD) | 1 (A) 20.76 | 2 (B) 16.72 | 3 (C) 19.36 | |
| 1 (A) 20.76 | | 4.04 (5.23) | 1.40 (5.23) | |
| 2 (B) 16.72 | 4.04 (5.23) | | 2.64 (5.23) | |
| 3 (C) 19.36 | 1.40 (5.23) | 2.64 (5.23) | | |

<Figure 9.2.5> Multiple comparisons by car

2) If this data have been extracted using the randomized block design, the block sum of squares will be separated from the error sum of squares. Adding Driver variable to 'by Group' in the variable selection box of 『eStat』 will give you a scatter plot of driver-specific fuel mileage for each car type as shown in <Figure 9.2.6>. This scatter plot shows a significant difference in fuel mileage per driver.

**Example 9.2.1 Answer (continued)**




<Figure 9.2.6> Fuel mileages for each driver

- Click the [ANOVA F-Test] button in the options window below the graph to reveal the two-way mean table shown in <Figure 9.2.7> and the ANOVA table shown in <Figure 9.2.8>. This ANOVA table clearly shows a decrease in error sum of squares and reduces significantly the mean squares of errors. This is due to the large variation between drivers being separated from the error variation. Factor B (driver) represents the block sum of squares separated from error term. The p-value shows that, the block (driver) effect is statistically significant. The $F$ value for the hypothesis $H_0$ : $\alpha_1 = \alpha_2 = \alpha_3 = 0$ of fuel mileage by Factor A (car type) is 43.447 and is greater than $F_{2,8,0.05}$= 4.46, so you can reject the $H_0$ at the significance level of 0.05. Consequently, significant differences in fuel mileages between car types can be found by removing the variation of the block in the error term.

| Two-dimension Statistics | | | | | | |
|---|---|---|---|---|---|---|
| Observation<br>Mean<br>Std Dev | Factor B (Driver) Level1 (1) | Factor B (Driver) Level2 (2) | Factor B (Driver) Level3 (3) | Factor B (Driver) Level4 (4) | Factor B (Driver) Level5 (5) | Factor A Level i Total |
| FactorA (Car) Level1 (A) | 1<br>22.400<br>NaN | 1<br>16.100<br>NaN | 1<br>19.700<br>NaN | 1<br>21.100<br>NaN | 1<br>24.500<br>NaN | 5<br>20.760<br>3.148 |
| FactorA (Car) Level2 (B) | 1<br>16.300<br>NaN | 1<br>12.600<br>NaN | 1<br>15.900<br>NaN | 1<br>17.800<br>NaN | 1<br>21.000<br>NaN | 5<br>16.720<br>3.054 |
| FactorA (Car) Level3 (C) | 1<br>20.200<br>NaN | 1<br>15.200<br>NaN | 1<br>18.700<br>NaN | 1<br>18.900<br>NaN | 1<br>23.800<br>NaN | 5<br>19.360<br>3.097 |
| Factor B Level j Total | 3<br>19.633<br>3.089 | 3<br>14.633<br>1.818 | 3<br>18.100<br>1.970 | 3<br>19.267<br>1.680 | 3<br>23.100<br>1.852 | 15<br>18.947<br>3.353 |
| Missing Observations | 0 | | | | | |

<Figure 9.2.7> Two-way mean table by car and driver
(There is no standard deviation of single data and denoted as NaN)

**Example 9.2.1 Answer (continued)**

| Analysis of Variance | | | | | |
|---|---|---|---|---|---|
| Factor | Sum of Squares | deg of freedom | Mean Squares | F value | p value |
| Factor A (Car) | 42.085 | 2 | 21.043 | 43.447 | < 0.0001 |
| Factor B (Driver) | 111.437 | 4 | 27.859 | 57.521 | < 0.0001 |
| Error | 3.875 | 8 | 0.484 | | |
| Total | 157.397 | 14 | | | |

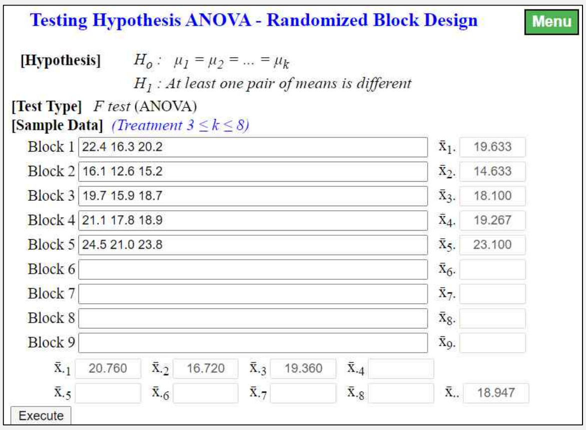
<Figure 9.2.8> ANOVA table for randomized block design

- In average, car type A has the best fuel mileage than other car types. In order to examine more about the differences between car types, the multiple comparisons test in the previous section can be applied. In this example, you can use one HSD value for all mean comparisons, because the number of repetitions at each level is the same.

$$HSD = q_{3,8\,;\,0.05}\sqrt{\frac{MSE}{r}} = (4.041)\sqrt{\frac{0.484}{5}} = 1.257$$

Therefore, there is a significant difference in fuel mileage between all three types of cars, since the differences between the mean values (4.04, 1.40, 2.64) are all greater than the critical value of 1.257.

- The same analysis of randomized design can be done using 『eStatU』 by following data input and clicking [Execute] button.




<Figure 9.2.9> Data input for 『eStatU』 RBD

**[Practice 9.2.1]**

The following is the result of an agronomist's survey of the yield of four varieties of wheat by using the randomized block design of the three cultivated areas (block). Test whether the mean yields of the four wheats are the same or not with 5% significance level.

| | | Cultivated Area | | |
|---|---|---|---|---|
| | | 1 | 2 | 3 |
| Wheat Type | A | 50 | 60 | 56 |
| | B | 59 | 52 | 51 |
| | C | 55 | 55 | 52 |
| | D | 58 | 58 | 55 |

Ex ⇨ eBook ⇨ PR090201_WheatAreaYield.csv

### 9.2.3 Latin Square Design

- In the experiments of randomized block design for measuring the fuel mileage explained in the previous section, there is one extraneous block variation which is the driver. If the researcher feels that there is an additional variation such as road type, there are two identifiable sources of extraneous block variations, i.e., two block variables. In this case, the researcher needs a design that will isolate and remove both sources of block variables from residual. The Latin square design is such a design.
- In the Latin square design, we assign one sources of extraneous variation to the columns of the square and the second source of extraneous variation to the rows of the square. We then assign the treatments in such a way that each treatment occurs one and only once in each row and each column. The number of rows, the number of columns, and the number of treatments, therefore, are all equal.
- Table 9.2.5 shows a 3 × 3 typical Latin squares with three rows, three columns and three treatments designated by capital letters A, B, C.

Table 9.2.5 Fuel mileage data by three drivers and three road types of three car types (A, B, C)

| | | Column 1<br>Road 1 | Column 2<br>Road 2 | Column 3<br>Road 3 |
|---|---|---|---|---|
| Row 1 | Driver 1 | A | B | C |
| Row 2 | Driver 2 | B | C | A |
| Row 3 | Driver 3 | C | A | B |

Table 9.2.6 shows a 4 × 4 typical Latin squares with four rows, four columns and four treatments designated by capital letters A, B, C, D.

Table 9.2.6 Fuel mileage data by four drivers and four road types of four car types (A, B, C, D)

| | | Column 1<br>Road 1 | Column 2<br>Road 2 | Column 3<br>Road 3 | Column 4<br>Road 4 |
|---|---|---|---|---|---|
| Row 1 | Driver 1 | A | B | C | D |
| Row 2 | Driver 2 | B | C | D | A |
| Row 3 | Driver 3 | C | D | A | B |
| Row 4 | Driver 4 | D | A | B | C |

- In the Latin square design, treatments can be assigned randomly in such a way that the car type occurs one and only once in each row and each column.. Therefore, there are many possible designs of 3 × 3 and 4 × 4 Latin square. We get randomization in the Latin square by randomly selection a square of the desired dimension from all possible squares of that dimension. One method of doing this is to randomly assign a different treatments to each cell in each column, with the restriction that each treatment must appear one and only once in each row.
- Small Latin squares provided only a small number of degrees of freedom for the error mean square. So a minimum size of 5 × 5 is usually recommended.
- The hypothesis of Latin square design with $r$ treatments is as follows:

| | |
|---|---|
| Null hypothesis | $H_0 : \mu_1 = \mu_2 = \cdots = \mu_r$ |
| Alternative hypothesis | $H_1$ : At least one pair of $\mu_i$ is not equal |

- Statistical model of the $r \times r$ Latin square design with $r$ treatments can be represented as follows:

$$Y_{ijk} = \mu + \alpha_i + \beta_j + \gamma_k + \epsilon_{ijk} \quad i=1,2,\cdots,r \quad j=1,2,\cdots,r \quad k=1,2,\cdots,r$$
where $\mu_i = \mu + \alpha_i$

In this equation, $\alpha_i$ is the effect of $i^{th}$ level of the row block variable to the response variable and $\beta_j$ is the effect of $j^{th}$ level of the column block variable to the response variable. $\gamma_k$ is the effect of $k^{th}$ level of the response variable.

- Notation for row averages, column averages and treatment averages of $r$ × $r$ Latin squre data are as follows;

Table 9.2.7 Notation for row means, column means and treatment averages of $r$ × $r$ Latin squre data

| | Column 1 | Column 2 | ⋯ Column r | Row Average |
|---|---|---|---|---|
| Row 1 | | | | $\overline{Y}_{1..}$ |
| Row 2 | | $Y_{ijk}$ | | $\overline{Y}_{2..}$ |
| ⋯ | | | | ⋯ |
| Row r | | | | $\overline{Y}_{r..}$ |
| Column Average | $\overline{Y}_{.1.}$ | $\overline{Y}_{.2.}$ | ⋯ $\overline{Y}_{.r.}$ | $\overline{Y}_{...}$ |

Treatment average: $\overline{Y}_{..1}$ $\overline{Y}_{..2}$ ⋯ $\overline{Y}_{..r}$

- In the Latin square design, the variation resulting from the difference between levels of two block variables can be separated from the error term of the variation of the factor independently. In the Latin square design, the total variation is divided into as follows:

$$Y_{ijk} - \overline{Y}_{...} = (Y_{ijk} - \overline{Y}_{i..} - \overline{Y}_{.j.} - \overline{Y}_{..k} + 2\overline{Y}_{...}) + (\overline{Y}_{i..} - \overline{Y}_{...}) + (\overline{Y}_{.j.} - \overline{Y}_{...}) + (\overline{Y}_{..k} - \overline{Y}_{...})$$

If you square both sides of the equation above and then combine for all $i,j,k$, you can obtain the following sums of squares:

Total sum of squares, degrees of freedom $r^3 - 1$

$$\text{SST} = \sum_{i=1}^{r}\sum_{i=1}^{r}\sum_{j=1}^{r}(Y_{ijk} - \overline{Y}_{...})^2$$

Error sum of squares, degrees of freedom $r^2 - 3r + 2$

$$\text{SSE} = \sum_{i=1}^{r}\sum_{j=1}^{r}\sum_{k=1}^{r}(Y_{ijk} - \overline{Y}_{i..} - \overline{Y}_{.j.} - \overline{Y}_{..k} + 2\overline{Y}_{...})^2$$

Row sum of squares, degrees of freedom $r - 1$

$$\text{SSR} = \sum_{i=1}^{r}\sum_{j=1}^{r}\sum_{k=1}^{r}(\overline{Y}_{i..} - \overline{Y}_{...})^2$$

Column sum of squares, degrees of freedom $r - 1$

$$\text{SSC} = \sum_{i=1}^{r}\sum_{j=1}^{r}\sum_{k=1}^{r}(\overline{Y}_{.j.} - \overline{Y}_{...})^2$$

Treatment sum of squares, degrees of freedom $r - 1$

$$\text{SSTr} = \sum_{i=1}^{r}\sum_{j=1}^{r}\sum_{k=1}^{r}(\overline{Y}_{..k} - \overline{Y}_{...})^2$$

- The following facts are always established in the Latin square design. Table 9.2.8 shows the ANOVA table of the Latin square design. In this ANOVA table, Division of the sum of squares and degrees of freedom

Sum of squares : $\quad \text{SST} \quad = \quad \text{SSE} + \text{SSR} + \text{SSC} + \text{SSTr}$

Degrees of freedom : $r^3-1 \quad = \quad r^2-3r+2 + (r-1) + (r-1) + (r-1)$

Table 9.2.8 ANOVA table of the Latin square design

| Variation | Sum of Squares | Degrees of freedom | Mean Squares | F value |
|---|---|---|---|---|
| Treatment | SSTr | $r-1$ | $\text{MSTr} = \dfrac{\text{SSTr}}{\text{r}-1}$ | $F_0 = \dfrac{MSTr}{MSE}$ |
| Row | SSR | $r-1$ | $\text{MSR} = \dfrac{\text{SSR}}{\text{r}-1}$ | |
| Column | SSC | $r-1$ | $\text{MSC} = \dfrac{\text{SSC}}{\text{r}-1}$ | |
| Error | SSE | $r^2-3r+2$ | $\text{MSE} = \dfrac{\text{SSE}}{\text{r}^2-3\text{r}+2}$ | |
| Total | SST | $r^3-1$ | | |

**Example 9.2.2**

Table 9.2.9 is the fuel mileage data of four car types (A, B, C, D) measured by four drivers and four road types with Latin square design.

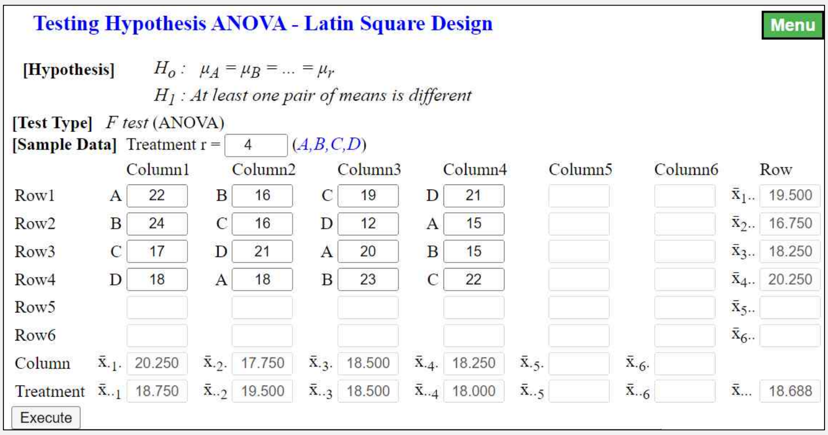
Table 9.2.9 Fuel mileage data by four drivers and four road types of four car types (A, B, C, D)

| | | Column 1 Road 1 | Column 2 Road 2 | Column 3 Road 3 | Column 4 Road 4 |
|---|---|---|---|---|---|
| Row 1 | Driver 1 | A(22) | B(16) | C(19) | D(21) |
| Row 2 | Driver 2 | B(24) | C(16) | D(12) | A(15) |
| Row 3 | Driver 3 | C(17) | D(21) | A(20) | B(15) |
| Row 4 | Driver 4 | D(18) | A(18) | B(23) | C(22) |

Use 『eStatU』 to do the analysis of variance whether the four car types have the same fuel mileage.

**Answer**

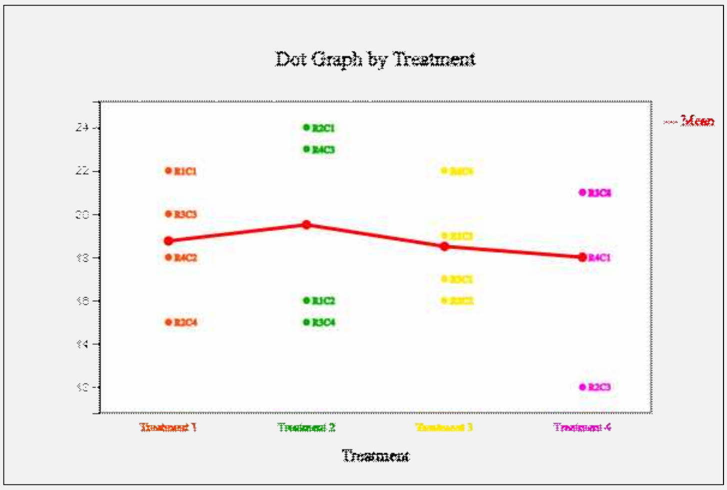
- In 『eStatU』 - 'Testing Hypothesis ANOVA – Latin Square Design', select the number of treatment r = 4 and enter data as shown in <Figure 9.2.10>.




<Figure 9.2.10> Data input for Latin square design in 『eStatU』

**Example 9.2.2 Answer (continued)**

- Click [Execute] button to show Dot graph by car type in Latin square design as <Figure 9.2.11> and ANOVA table as in <Figure 9.2.12>. The dot graph and result of the ANOVA is that there is no difference in fuel mileage between the car types.




<Figure 9.2.11> Dot graph by car type in Latin square design

**Analysis of Variance**

| Factor | Sum of Squares | deg of freedom | Mean Squares | F value | p value |
|---|---|---|---|---|---|
| Treatment | 4.688 | 3 | 1.563 | 0.075 | 0.9710 |
| Row Var | 28.188 | 3 | 9.396 | | |
| Col Var | 14.188 | 3 | 4.729 | | |
| Error | 124.375 | 6 | 20.729 | | |
| Total | 171.438 | 15 | | | |

<Figure 9.2.12> ANOVA table of Latin square design

**[Practice 9.2.2]**

To study the effect of packaging on the sales of a certain cereal, a researcher tries three different packaging methods (treatments) at four different times of the week (columns) in four different supermarket chains (rows). The variable of interest is daily salse. The following table shows the results of the study. Do these data show a significant difference in shoppers' response to the different packaging methods? Let $\alpha$ = 0.05.

| | | Time of week | | | |
|---|---|---|---|---|---|
| | | 1 | 2 | 3 | 4 |
| Store | 1 | A(50) | B(60) | C(56) | D(63) |
| | 2 | B(59) | C(52) | D(51) | A(57) |
| | 3 | C(55) | D(55) | A(52) | B(56) |
| | 4 | D(58) | A(58) | B(55) | C(61) |

## 9.3 Analysis of Variance for Experiments of Two Factors

- If there are two factors affecting the response variable, the analysis is called a two-way analysis of variances. This technique is frequently used in experiments such as engineering, medicine and agriculture. The response variable is observed at each combination of levels of two factors (denoted as A and B). In general, it is advisable to repeat at least two experiments at each combination of levels of two factors, if possible, in order to increase the reliability of the experimental results.
- When data are obtained from repeated experiments at each factor level, the two-way ANOVA tests whether the population means of each level of factor A are the same (called the **main effect** test of the factor A) as the one-way ANOVA, or tests whether the population means of each level of factor B are the same (called the main effect test of the factor B). In addition, the two-way ANOVA tests whether the effect of one factor A is influenced by each level of the other factor B (called the **interaction effect** test). For example, in a chemical process, if the higher the pressure when the temperature is low, the greater the amount of products, and the lower the pressure when the temperature is high, the greater the amount of products, the interaction effect exists between the two factors of temperature and pressure. The interaction effect exists where the effects of one factor change with changes in the level of another factor.

**Definition**

**Main effect and Interaction effect**

When data are obtained from repeated experiments at each factor level, the two-way ANOVA tests whether the population means of each level of factor A (called the **main effect** test of the factor A) are the same as the one-way ANOVA, or tests whether the population means of each level of factor B are the same (called the main effect test of the factor B).

The two-way ANOVA also tests whether the effect of one factor A is influenced by each level of the other factor B (called the **interaction effect** test).

**Example 9.3.1**

Table 9.3.1 shows the yield data of three repeated agricultural experiments for each combination of four fertilizer levels and three rice types to investigate the yield of rice.

Table 9.3.1 Yield of rice by fertilizers and types of rice (unit kg)

| Fertilizer | Types of rice | | |
|---|---|---|---|
| | 1 | 2 | 3 |
| 1 | 64,66,70 | 72,81,64 | 74,51,65 |
| 2 | 65,63,58 | 57,43,52 | 47,58,67 |
| 3 | 59,68,65 | 66,71,59 | 58,45,42 |
| 4 | 58,50,49 | 57,61,53 | 53,59,38 |

Ex ⇨ eBook ⇨ EX090301_YieldByRiceFertilzer.csv

1) Find the average yield for each combination of fertilizers and rice types.
2) Using 『eStat』, draw a scatter plot with the rice types (1, 2 and 3) as X-axis and the yield as Y-axis. Separate the color of dots in the scatter plot by the type of fertilizer. Then, show the average of the combinations at each level on the scatter plot and connect them with lines for each type of fertilizer to observe.
3) Test the main effects of fertilizers and rice types and test the interaction effect of the two factors.
4) Using 『eStat』, check the result of the two-way analysis of variance.

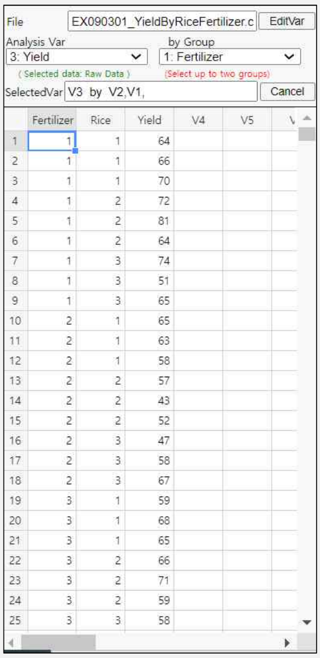
**Example 9.3.1 Answer**

1) For convenience, let us call the fertilizer as the factor A and the rice type as factor B. The averages of the rice yield for each level combination of two factors are shown in Table 9.3.2. Denote the $k^{th}$ rice yield, $y_{ijk}$, and average $\overline{y}_{ij\cdot}$ of each combination of $j^{th}$ level of factor A and $i^{th}$ level of factor B. Also, denote the average of $j^{th}$ level of factor A as $\overline{y}_{\cdot j \cdot}$, the average of $i^{th}$ level of factor B as $\overline{y}_{i \cdot\cdot}$, and the global average as $\overline{y}_{\cdot\cdot\cdot}$.

Table 9.3.2 Average yield of rice by fertilizers and types of rice (unit kg)

| Fertilizer (Factor B) | Types of Rice (Factor A) | | | Row Average |
|---|---|---|---|---|
| | 1 | 2 | 3 | |
| 1 | $\overline{y}_{11\cdot} = 66.7$ | $\overline{y}_{12\cdot} = 72.3$ | $\overline{y}_{13\cdot} = 63.3$ | $\overline{y}_{1\cdot\cdot} = 67.4$ |
| 2 | $\overline{y}_{21\cdot} = 62.0$ | $\overline{y}_{22\cdot} = 50.7$ | $\overline{y}_{23\cdot} = 57.3$ | $\overline{y}_{2\cdot\cdot} = 56.7$ |
| 3 | $\overline{y}_{31\cdot} = 64.0$ | $\overline{y}_{32\cdot} = 65.3$ | $\overline{y}_{33\cdot} = 48.3$ | $\overline{y}_{3\cdot\cdot} = 59.2$ |
| 4 | $\overline{y}_{41\cdot} = 52.3$ | $\overline{y}_{42\cdot} = 57.0$ | $\overline{y}_{43\cdot} = 50.0$ | $\overline{y}_{4\cdot\cdot} = 53.1$ |
| Column Average | $\overline{y}_{\cdot 1\cdot} = 61.3$ | $\overline{y}_{\cdot 2\cdot} = 61.3$ | $\overline{y}_{\cdot 3\cdot} = 54.8$ | $\overline{y}_{\cdot\cdot\cdot} = 59.1$ |

2) To draw a scatter plot for the two-way ANOVA using 『eStat』, enter data as <Figure 9.3.1> where the fertilizer is variable 1, the rice type is variable 2 and the rice yield is variable 3.

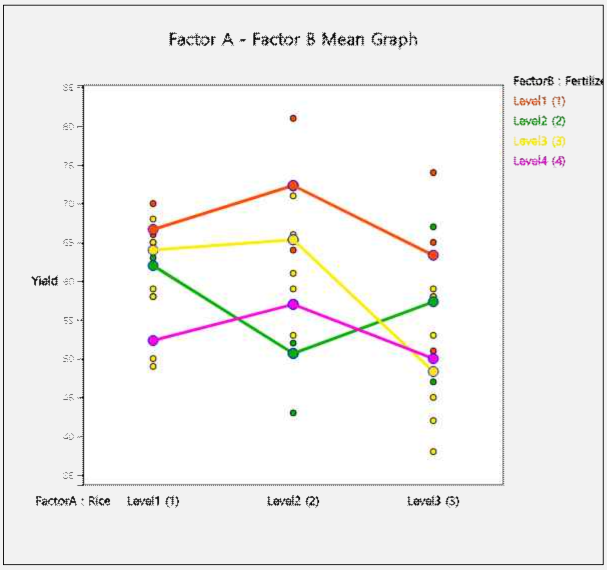



| | Fertilizer | Rice | Yield | V4 | V5 |
|---|---|---|---|---|---|
| 1 | 1 | 1 | 64 | | |
| 2 | 1 | 1 | 66 | | |
| 3 | 1 | 1 | 70 | | |
| 4 | 1 | 2 | 72 | | |
| 5 | 1 | 2 | 81 | | |
| 6 | 1 | 2 | 64 | | |
| 7 | 1 | 3 | 74 | | |
| 8 | 1 | 3 | 51 | | |
| 9 | 1 | 3 | 65 | | |
| 10 | 2 | 1 | 65 | | |
| 11 | 2 | 1 | 63 | | |
| 12 | 2 | 1 | 58 | | |
| 13 | 2 | 2 | 57 | | |
| 14 | 2 | 2 | 43 | | |
| 15 | 2 | 2 | 52 | | |
| 16 | 2 | 3 | 47 | | |
| 17 | 2 | 3 | 58 | | |
| 18 | 2 | 3 | 67 | | |
| 19 | 3 | 1 | 59 | | |
| 20 | 3 | 1 | 68 | | |
| 21 | 3 | 1 | 65 | | |
| 22 | 3 | 2 | 66 | | |
| 23 | 3 | 2 | 71 | | |
| 24 | 3 | 2 | 59 | | |
| 25 | 3 | 3 | 58 | | |

<Figure 9.3.1> Data input for two-way ANOVA in 『eStat』

- In the variable selection box which appears by clicking the ANOVA icon on the main menu, select 'Analysis Var' as Yield and 'By Group' as Rice and Fertilizer, then the scatter plot of the yield by rice type will appear as in <Figure 9.3.2>. In addition, the average yields at each rice type by fertilizer are marked as dots linking them with lines by fertilizer. In this graph, rice type 1 always yields more than rice type 3 regardless of the fertilizer used. Rice type 2 varies in yield depending on the type of fertilizer used, which shows the existence of interaction, and the use of fertilizer 1 usually results in a high yield regardless of the rice types.

**Example 9.3.1 Answer (continued)**




<Figure 9.3.2> Yields by rice types and fertilizer types

3) Testing the factor A, which is to test the main effect of rice types, implies to test the following null hypothesis.

$H_0$: The average yields of the three rice types are the same.

- If the null hypothesis is rejected, we conclude that the main effect of rice types exists. In order to test the main effect of rice types, as in the one-way analysis of variance, the sum of squared distances from each average yield $\overline{y}_{\cdot j \cdot}$ of rice type $j$ to the overall average yield $\overline{y}_{\cdots}$.

$$SSA = 12(61.3 - \overline{y}_{\cdots})^2 + 12(61.3 - \overline{y}_{\cdots})^2 + 12(54.8 - \overline{y}_{\cdots})^2 = 342.39$$

where the weight of 12 of each sum of squares is the number of data for each rice type. Since there are 3 rice types, the degrees of freedom of $SSA$ is (3-1) and we call the sum of squares $SSA$ divided by (3-1), $SSA/(3-1)$, is the mean squares of factor A, $MSA$.

- Testing the factor B, which is to test the main effect of fertilizer types, implies to test the following null hypothesis.

$H_0$: The average yields of the four fertilizer types are the same.

- If the null hypothesis is rejected, we conclude that the main effect of fertilizer types exists. In order to test the main effect of fertilizer types, as in the one-way analysis of variance, the sum of squared distances from each average yield $\overline{y}_{i\cdot\cdot}$ of fertilizer type $i$ to the overall average yield $\overline{y}_{\cdots}$,

$$SSB = 9(67.4 - \overline{y}_{\cdots})^2 + 9(56.7 - \overline{y}_{\cdots})^2 + 9(59.2 - \overline{y}_{\cdots})^2 + 9(53.1 - \overline{y}_{\cdots})^2 = 1002.89$$

where the weight of 9 of each sum of squares is the number of data for each fertilizer type. Since there are 4 fertilizer types, the degrees of freedom of $SSB$ is (4-1) and we call the sum of squares $SSB$ divided by (4-1), $SSB/(4-1)$, is the mean squares of factor B, $MSB$.

- Testing the interaction effect of rice and fertilizer (represented as factor AB) is to test the following null hypothesis.

$H_0$: There is no interaction effect between rice type and fertilizer type.

**Example 9.3.1 Answer (continued)**

- If the null hypothesis is rejected, we conclude that there is an interaction effect between rice types and fertilizer types. In order to test the interaction effect, the sum of squared distances from each average yield $\overline{y}_{ij.}$ subtracting the average yield $\overline{y}_{i..}$ of fertilizer type $i$, subtracting the average yield $\overline{y}_{.j.}$ of rice type $j$, adding the overall average yield $\overline{y}_{...}$.

$$
\begin{aligned}
\text{SSAB} = \; & 3(66.7-\overline{y}_{1..}-\overline{y}_{.1.}+\overline{y}_{...})^2+3(72.3-\overline{y}_{1..}-\overline{y}_{.2.}+\overline{y}_{...})^2 \\
& +3(63.3-\overline{y}_{1..}-\overline{y}_{.3.}+\overline{y}_{...})^2+3(62.0-\overline{y}_{2..}-\overline{y}_{.1.}+\overline{y}_{...})^2 \\
& +3(50.7-\overline{y}_{2..}-\overline{y}_{.2.}+\overline{y}_{...})^2+3(57.3-\overline{y}_{2..}-\overline{y}_{.3.}+\overline{y}_{...})^2 \\
& +3(64.0-\overline{y}_{3..}-\overline{y}_{.1.}+\overline{y}_{...})^2+3(65.3-\overline{y}_{3..}-\overline{y}_{.2.}+\overline{y}_{...})^2 \\
& +3(48.3-\overline{y}_{3..}-\overline{y}_{.3.}+\overline{y}_{...})^2+3(52.3-\overline{y}_{4..}-\overline{y}_{.1.}+\overline{y}_{...})^2 \\
& +3(57.0-\overline{y}_{4..}-\overline{y}_{.2.}+\overline{y}_{...})^2+3(50.0-\overline{y}_{4..}-\overline{y}_{.3.}+\overline{y}_{...})^2 \\
= \; & 588.94
\end{aligned}
$$

where the weight of 3 of each sum of squares is the number of data for each cell of rice and fertilizer type. The degrees of freedom of $SSAB$ is (3-1)(4-1) and we call the sum of squares $SSAB$ divided by (3-1)(4-1), $SSAB/((3-1)(4-1))$ is the mean squares of interaction AB, $MSAB$.

- It is not possible to test each effect immediately using these sum of squares, but the error sum of squares should be calculated. In order to calculate the error sum of squares, first we calculate the total sum of squares which is the sum of the squared distances from each data to the overall average.

$$
\begin{aligned}
\text{SST} = \; & (64-\overline{y}_{...})^2+(66-\overline{y}_{...})^2+(70-\overline{y}_{...})^2 \\
& + \;...\; +(53-\overline{y}_{...})^2+(59-\overline{y}_{...})^2+(38-\overline{y}_{...})^2 = 3267.56
\end{aligned}
$$

This total sum of squares can be proven mathematically to be the sum of the other sums of squares as follows:

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

Therefore, the error sum of squares can be calculated as follows:

$$\text{SSE} = \text{SST} - (\text{SSA} + \text{SSB} + \text{SSAB}) = 1333.33$$

- If the yields on each rice type or fertilizer type are assumed to be normal and the variances are the same, the statistic which divides the each mean squares by the error mean squares follows $F$ distribution. Therefore, the main effects and interaction effect can be tested using $F$ distributions. If the interaction effect is separated, we test them first. Testing results using the 5% significance level are as follows:

① Testing of the interaction effect on rice and fertilizer:

$$F_0 = \frac{MSAB}{MSE} = \frac{\frac{\text{SSAB}}{(3-1)(4-1)}}{\frac{\text{SSE}}{24}} = 1.77$$

$$F_{6,24\,;\,0.05} = 2.51$$

Since $F_0 < F_{6,24\,;\,0.05}$, we conclude that there is no interaction. The interaction on rice and fertilizer in <Figure 9.3.2> is so small which is not statistically significant and it may due to other kind of random error. The calculated p-value of $F_0 = 1.77$ using『eStat』is 0.1488.

**Example 9.3.1 Answer (continued)**

② Testing of the main effect on rice types (Factor A):

$$F_0 \;=\; \frac{MSA}{MSE} \;=\; \frac{\frac{\mathrm{SSA}}{(3-1)}}{\frac{\mathrm{SSE}}{24}} \;=\; 3.08$$

$$F_{2,24\,;\,0.05} \;=\; 3.40$$

Since $F_0$ < $F_{2,24\,;\,0.05}$, we can not reject the null hypothesis that average yields of rice types are the same. There is not enough evidence statistically that average yields are different depending on rice types. The calculated p-value of $F_0 = 3.08$ using 『eStat』 is 0.0644.

③ Testing of the main effect on fertilizer types (Factor B):

$$F_0 \;=\; \frac{MSAB}{MSE} \;=\; \frac{\frac{\mathrm{SSB}}{(4-1)}}{\frac{\mathrm{SSE}}{24}} \;=\; 6.02$$

$$F_{3,24\,;\,0.05} \;=\; 3.01$$

Since $F_0$ > $F_{3,24\,;\,0.05}$, we reject the null hypothesis that average yields of fertilizer types are the same. There is enough statistical evidence which shows that average yields are different depending on fertilizer types. Since there is no interaction effect by 1), we can conclude that fertilizer 1 produces more yields than other fertilizer. The calculated p-value of $F_0 = 6.02$ using 『eStat』 is 0.0033.

- The result of the two-way analysis of variances is as Table 9.3.3.

Table 9.3.3 two-way analysis of variance of yields by rice and fertilizer types

| Factor | Sum of Squares | degrees of freedom | Mean Squares | F value | $p$-value |
|---|---|---|---|---|---|
| Rice Type | 342.3889 | 2 | 171.1944 | 3.0815 | 0.0644 |
| Feritlizer Type | 1002.8889 | 3 | 334.2963 | 6.0173 | 0.0033 |
| Interaction | 588.9444 | 6 | 98.1574 | 1.7668 | 0.1488 |
| Error | 1333.3333 | 24 | 55.5556 | | |
| Total | 3267.5556 | 35 | | | |

4) If you press the [ANOVA F-test] button in the options window below <Figure 9.3.2> of 『eStat』, the two-dimensional table of means / standard deviations for each level combination as in <Figure 9.3.3> and the two-way analysis of variance table as in <Figure 9.3.4> will appear in the Log Area.

**Two-dimension Statistics**

| Observation<br>Mean<br>Std Dev | Factor B (Fertilizer) Level1 (1) | Factor B (Fertilizer) Level2 (2) | Factor B (Fertilizer) Level3 (3) | Factor B (Fertilizer) Level4 (4) | Factor A Level i Total |
|---|---|---|---|---|---|
| FactorA (Rice) Level1 (1) | 3<br>66.667<br>3.055 | 3<br>62.000<br>3.606 | 3<br>64.000<br>4.583 | 3<br>52.333<br>4.933 | 12<br>61.250<br>6.649 |
| FactorA (Rice) Level2 (2) | 3<br>72.333<br>8.505 | 3<br>50.667<br>7.095 | 3<br>65.333<br>6.028 | 3<br>57.000<br>4.000 | 12<br>61.333<br>10.263 |
| FactorA (Rice) Level3 (3) | 3<br>63.333<br>11.590 | 3<br>57.333<br>10.017 | 3<br>48.333<br>8.505 | 3<br>50.000<br>10.817 | 12<br>54.750<br>10.788 |
| Factor B Level j Total | 9<br>67.444<br>8.338 | 9<br>56.667<br>8.078 | 9<br>59.222<br>9.972 | 9<br>53.111<br>6.990 | 36<br>59.111<br>9.662 |
| Missing Observations | 0 | | | | |

<Figure 9.3.3> Two dimensional mean / standard deviation table

**Example 9.3.1 Answer (continued)**

| Analysis of Variance | | | | | |
|---|---|---|---|---|---|
| Factor | Sum of Squares | deg of freedom | Mean Squares | F value | p value |
| Factor A (Rice) | 342.389 | 2 | 171.194 | 3.082 | 0.0644 |
| Factor B (Fertilizer) | 1002.889 | 3 | 334.296 | 6.017 | 0.0033 |
| Interaction | 588.944 | 6 | 98.157 | 1.767 | 0.1488 |
| Error | 1333.333 | 24 | 55.556 | | |
| Total | 3267.556 | 35 | | | |

<Figure 9.3.4> two-way analysis of variance table

- Let's generalize the theory of the two-way analysis of variance discussed in the example above. Let $Y_{ijk}$ be the random variable representing the $k^{th}$ observation at the $i^{th}$ level of factor A, which has $a$ number of levels, and $j^{th}$ level of factor B, which has $b$ number of levels. A statistical model of the two-way analysis of variances is as follows:

$$Y_{ijk} = \mu + \alpha_i + \beta_j + \gamma_{ij} + \epsilon_{ijk}, \quad i=1,2,\cdots,a\,;\, j=1,2,\cdots,b\,;\, k=1,2,\cdots,r$$

$\mu$ : total mean
$\alpha_i$ : effect of $i^{th}$ level of factor A
$\beta_j$ : effect of $j^{th}$ level of factor B
$\gamma_{ij}$ : interaction effect of $i^{th}$ level of factor A and $j^{th}$ level of factor B
$\epsilon_{ijk}$ : error terms which are independent and follow N(0,$\sigma^2$).

Assume that experiments are repeated $r$ times equally at the $i^{th}$ level of factor A and $j^{th}$ level of factor B. Therefore, the total number of observations is $n = abr$.

- The total sum of squared distances from each observation to the total mean $\overline{Y}_{...}$ can be partitioned as following sum of squares similar to the one-way analysis of variance.

Total sum of squares: $\mathrm{SST} = \sum_{i=1}^{a}\sum_{j=1}^{b}\sum_{k=1}^{r}(\mathrm{Y}_{ijk} - \overline{\mathrm{Y}}_{...})^2$ : degrees of freedom: $n-1$

Fator A sum of squares: $\mathrm{SSA} = \mathrm{br}\sum_{i=1}^{a}(\overline{\mathrm{Y}}_{i..} - \overline{\mathrm{Y}}_{...})^2$ : degrees of freedom: $a-1$

Factor B sum of squares: $\mathrm{SSB} = \mathrm{ar}\sum_{j=1}^{b}(\overline{\mathrm{Y}}_{.j.} - \overline{\mathrm{Y}}_{...})^2$ : degrees of freedom: $b-1$

Interaction sum of squares: $\mathrm{SSAB} = \mathrm{r}\sum_{i=1}^{a}\sum_{j=1}^{b}(\overline{\mathrm{Y}}_{ij.} - \overline{\mathrm{Y}}_{i..} - \overline{\mathrm{Y}}_{.j.} + \overline{\mathrm{Y}}_{...})^2$ : degrees of freedom: $(a-1)(b-1)$

Error sum of squares: $\mathrm{SSE} = \sum_{i=1}^{a}\sum_{j=1}^{b}\sum_{k=1}^{r}(\mathrm{Y}_{ijk} - \overline{\mathrm{Y}}_{ij.})^2$ : degrees of freedom: $n-ab$

**Partition of Sum of Squares and degrees of freedom**

Sum of Squares: $SST = SSA + SSB + SSAB + SSE$
degrees of freedom: $(n-1) = (a-1) + (b-1) + (a-1)(b-1) + (n-ab)$

- The two-way analysis of variance is summarized as Table 9.3.4.

Table 9.3.4 two-way analysis of variance table

| Factor | Sum of Squares | Degree of Freedom | Mean Squares | F value |
|---|---|---|---|---|
| Factor A | SSA | $a-1$ | MSA = SSA/$(a-1)$ | $F_1$ = MSA/MSE |
| Factor B | SSB | $b-1$ | MSB = SSB/$(b-1)$ | $F_2$ = MSB/MSE |
| Interaction | SSAB | $(a-1)(b-1)$ | MSAB = SSAB/$((a-1)(b-1))$ | $F_3$ = MSAB/MSE |
| Error | SSE | $n-ab$ | MSE = SSE/$(n-ab)$ | |
| Total | SST | $n-1$ | | |

**Two-way analysis of variance without repetition of experiments**

If there is no repeated observation at each level combination of two factors, the interaction effect can not be estimated and the row of interaction factor is deleted from the above two-way ANOVA table. In this case, the analysis of variance table is the same as the randomized block design as Table 9.2.3.

- Testing hypothesis for the main effects and interaction effect of factor A and factor B are as follows. If the interaction effect is separated, it is reasonable to test the interaction effect first. This is because, depending on the significance of the interaction effect, the method of interpreting the result of the main effect test of each factor can be different.

1) $F$ Test for the interaction effect:

$H_0 : \gamma_{ij} = 0,\ i=1,2,\cdots,a;\ j=1,2,\cdots,b$
If $F_3 = \mathrm{MSAB}/\mathrm{MSE}$ > $F_{(a-1)(b-1),\, n-ab\, ;\, \alpha}$, then reject $H_0$

2) $F$ Test for the main effect of factor A:

$H_0 : \alpha_1 = \alpha_2 = \cdots = \alpha_a = 0$
If $F_1 = \mathrm{MSA}/\mathrm{MSE}$ > $F_{(a-1),\, n-ab\, ;\, \alpha}$, then reject $H_0$

3) $F$ Test for the main effect of factor B:

$H_0 : \beta_1 = \beta_2 = \cdots = \beta_b = 0$
If $F_2 = \mathrm{MSB}/\mathrm{MSE}$ > $F_{b-1,\, n-ab\, ;\, \alpha}$, then reject $H_0$

( 『eStat』 calculates the p-value for each of these tests and tests them using it. That is, for each test, if the p-value is less than the significance level, the null hypothesis $H_0$ is rejected.)

- If the test for interaction effect is not significant, a test of the main effects of each factor can be performed to test significant differences between levels. However, if there is a significant interaction effect, the test for the main effects of each factor is meaningless, so an analysis should be made on which level combinations of factors show differences in the means.

- If you conclude that significant differences between the levels of a factor as in the one-way analysis of variance exist there, you can compare confidence intervals at each level to see which level of the differences appears. And a residual analysis is necessary to investigate the validity of the assumption.

**[Practice 9.3.1]**

The result of an experiment at a production plant of an electronic component to investigate the life of the product due to changes in temperature $(T_1, T_2)$ and humidity $(O_1, O_2)$ is as follows. Analyze data using the analysis of variance with 5% significance level.

| (Unit: Time) | $O_1$ | $O_2$ |
|---|---|---|
| $T_1$ | 6.29 | 5.95 |
| | 6.38 | 6.05 |
| | 6.25 | 5.89 |
| $T_2$ | 5.80 | 6.32 |
| | 5.92 | 6.44 |
| | 5.78 | 6.29 |

Ex ⇨ eBook ⇨ PR090301_LifeByTemperatureHumidity.csv

### Design of experiments for the two-way analysis of variances

Even in the two-way analysis of variance, obtaining sample data at each level of two factors in engineering or in agriculture can be influenced by other factors and should be careful in sampling. In order to accurately identify the differences that may exist between each level of a factor, it is advisable to make as few as possible influences from other factors. One of the most commonly used methods of doing this is completely randomized design which makes the entire experiments random. There are many other experimental design methods, and for more information, refer to the references to the experimental design of several factors.

# Exercise

9.1 Complete the following ANOVA table.

| Factor | $SS$ | df | $MS$ | $F$ ratio |
|---|---|---|---|---|
| Treatment | 154.9199 | 4 | ______ | ______ |
| Error | ________ | __ | | |
| Total | 200.4773 | 39 | | |

9.2 Answer the following questions based on this ANOVA table.

| Factor | $SS$ | df | $MS$ | $F$ ratio |
|---|---|---|---|---|
| Treatment | 5.05835 | 2 | 2.52917 | 1.0438 |
| Error | 65.42090 | 27 | 2.4230 | |

1) How many levels of treatment are compared?
2) How many total number of observations are there?
3) Can you conclude that the levels of treatment are significantly different with the 5% significance level? Why?

9.3 In order to test customers' responses to new products, four different exhibition methods (A, B, C and D) were used by a company. Each exhibition method was used in nine stores by selecting 36 stores that met the company's criteria. The total sales for the weekend are shown in the following table.

| Exhibition Method | Sales for the weekend in 9 stones (unit: 1000USD) | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| A | 5 | 6 | 7 | 7 | 8 | 6 | 7 | 7 | 6 |
| B | 2 | 2 | 2 | 3 | 3 | 2 | 3 | 3 | 2 |
| C | 2 | 2 | 3 | 3 | 2 | 2 | 2 | 3 | 3 |
| D | 6 | 6 | 7 | 8 | 8 | 8 | 6 | 6 | 6 |

1) Draw a scatter plot of sales (y axis) and exhibition method (x axis). Mark the average sales of each exhibition method and connect them with a line.
2) Test that the sales by each exhibition method are different in the amount of sales with the 5% significance level. Can you conclude that one of the exhibition methods shows significant effect on sales?

9.4 The following table shows mileages in km per liter obtained from experiments to compare three brands of gasoline. In this experiment, seven cars of the same type were used in a similar situation to reduce the variation of the car.

| Gasoline | mileage in km / liter | | | | | | |
|---|---|---|---|---|---|---|---|
| A | 14 | 19 | 19 | 16 | 15 | 17 | 20 |
| B | 20 | 21 | 18 | 20 | 19 | 19 | 18 |
| C | 20 | 26 | 23 | 24 | 23 | 25 | 23 |

1) Calculate the average mileages of each gasoline brand. Draw a scatter plot of gas milage (y axis) and gasoline brand (x axis) to compare.
2) From this data, test whether there are differences between gasoline brands for gas

milage with the 5% significance level.

9.5 The result of a survey on job satisfaction of three companies (A, B, and C) is as follows. Test whether the averages of job satisfaction of the three companies are different with the 5% significance level.

| Company | Job satisfaction score |
|---|---|
| A | 69 67 65 59 68 61 66 |
| B | 56 63 55 59 52 57 |
| C | 71 72 70 68 74 |

9.6 Psychologists were asked to investigate the job satisfaction of salespeople in three companies: A, B and C. Ten salespeople were randomly selected from each company and a test to measure the job satisfaction was conducted. Test scores are as follows. From this data, can we claim that the average scores of the job satisfaction of three companies are different with the significance level of 0.05?

| Company | Job satisfaction score | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| A | 67 | 65 | 59 | 59 | 58 | 61 | 66 | 53 | 51 | 64 |
| B | 66 | 68 | 55 | 59 | 61 | 66 | 62 | 65 | 64 | 74 |
| C | 87 | 80 | 67 | 89 | 80 | 84 | 78 | 65 | 72 | 85 |

9.7 An advertising agency experimented to find out the effects of various forms (A, B, C, D and E) of TV advertising. Fifty television viewers were shown five forms of TV commercials for a cold medicine in random order one by one. The effect of advertising after viewing was measured and recorded as follows. Test an appropriate hypothesis with the 5% significance level.

| Forms of TV Advertising | | | | |
|---|---|---|---|---|
| A | B | C | D | E |
| 20 23 21 | 28 27 22 | 33 34 25 | 33 29 31 | 49 41 41 |
| 23 26 24 | 28 23 29 | 26 27 33 | 29 27 25 | 39 41 48 |
| 26 23 20 | 27 25 28 | 25 32 25 | 26 26 33 | 43 43 46 |
| 24 | 21 | 34 | 32 | 35 |

9.8 The following is the result of an agronomist's survey of the yield of four varieties of wheat by using the randomized block design of three cultivated areas (block). Test whether the mean yields of the four wheats are the same or not with the 5% significance level.

| Wheat Type | Cultivated Area | | | Average ($\bar{y}_{i.}$) |
|---|---|---|---|---|
| | 1 | 2 | 3 | |
| A | 60 | 61 | 56 | 59 |
| B | 59 | 52 | 51 | 54 |
| C | 55 | 55 | 52 | 54 |
| D | 58 | 58 | 55 | 57 |

9.9 Answer the following questions based on the following ANOVA table.

| Factor | $SS$ | df | $MS$ | $F$ value | $p$-value |
|---|---|---|---|---|---|
| A | 12.3152 | 2 | 6.1575 | 29.4021 | $<$ 0.005 |
| B | 19.7844 | 3 | 6.5948 | 31.4898 | $<$ 0.005 |
| AB | 8.9416 | 6 | 1.4902 | 7.1159 | $<$ 0.005 |
| Error | 10.0525 | 48 | 0.2094 | | |
| Total | 51.0938 | 59 | | | |

1) What method of analysis was used?
2) What conclusions can be obtained from the above analysis table? The significance level is 0.05.

9.10 Research was conducted to compare the job satisfaction of workers in the assembly process with different working conditions. Another concern is the relationship between the job satisfaction and years of service. Observers would like to investigate the interaction effect between the years of service and working conditions. The following table shows the level of the job satisfaction obtained from the survey. Analyze the data using an appropriate methodology.

| | Working condition | | |
|---|---|---|---|
| Years of service | Good | Fair | Bad |
| < 5 | 12 | 10 | 8 |
| | 15 | 10 | 7 |
| | 15 | 9 | 7 |
| | 14 | 10 | 8 |
| | 12 | 9 | 6 |
| 5 - 10 | 12 | 10 | 10 |
| | 14 | 10 | 11 |
| | 12 | 14 | 12 |
| | 10 | 14 | 10 |
| | 11 | 10 | 14 |
| 11 or more | 9 | 10 | 12 |
| | 10 | 11 | 14 |
| | 9 | 10 | 15 |
| | 9 | 10 | 15 |
| | 10 | 12 | 15 |

9.11 The following table shows the degree of stress in the work and the level of anxiety among 27 workers classified as years of service. Analyze data using the analysis of variance with the 5% significance level.

| Factor A | Job-induced pressure (Factor B) | | |
|---|---|---|---|
| Years of service | Good | Fair | Bad |
| < 5 | 25 | 18 | 17 |
| | 28 | 23 | 24 |
| | 22 | 19 | 19 |
| 5 - 10 | 28 | 16 | 18 |
| | 32 | 24 | 22 |
| | 30 | 20 | 20 |
| 11 or more | 25 | 14 | 10 |
| | 35 | 16 | 8 |
| | 30 | 15 | 12 |

9.12 A fertilizer manufacturer hired a research team to study the yields of three grain seeds (A, B, C) and three types of fertilizer (1, 2, 3). Three grain seeds in combination of three types of fertilizer were used and the experiment were repeated three times at each combination of treatments. Each combination of treatments was randomly assigned to 27 different regions. Analyze data using the analysis of variance with the 5% significance level.

| Seed type | Fertilizer type 1 | 2 | 3 |
|---|---|---|---|
| A | 5 | 8 | 10 |
| | 8 | 8 | 9 |
| | 7 | 10 | 10 |
| B | 6 | 10 | 15 |
| | 8 | 12 | 14 |
| | 6 | 11 | 14 |
| C | 7 | 12 | 16 |
| | 8 | 12 | 10 |
| | 10 | 14 | 18 |

9.13 The result of an experiment at a production plant of an electronic component to investigate the life of the product due to changes in temperature ($T_1, T_2$) and humidity ($O_1, O_2$) is as follows. Analyze data using the analysis of variance with the 5% significance level.

| (Unit: Time) | $O_1$ | $O_2$ |
|---|---|---|
| $T_1$ | 6.29 | 5.95 |
| | 6.38 | 6.05 |
| | 6.25 | 5.89 |
| $T_2$ | 5.80 | 6.32 |
| | 5.92 | 6.44 |
| | 5.78 | 6.29 |

9.14 The result of a fertilizer manufacturer's experiment with the production of soybeans on two seeds using three types of fertilizer (A, B, and C) is as follows. Each fertilizer and seed were tested four times. Analyze data using the analysis of variance with the 5% significance level.

| | Fertilizer A | B | C |
|---|---|---|---|
| Seed 1 | 5 | 8 | 10 |
| | 8 | 8 | 12 |
| | 7 | 10 | 10 |
| | 6 | 10 | 10 |
| Seed 2 | 8 | 12 | 14 |
| | 6 | 11 | 16 |
| | 8 | 12 | 16 |
| | 10 | 14 | 18 |

# Multiple Choice Exercise

9.1 Who first announced the ANOVA method?

① Laspeyres ② Paasche
③ Fisher ④ Edgeworth

9.2 What are the abbreviation of the analysis of variance?

① ANOVA ② $t$
③ $F$ ④ $\chi^2$

9.3 Which areas are not the area of application for the analysis of variance?

① marketing survey ② quality control
③ economy forecasting ④ medical experiment

9.4 Which sampling distribution is used for the analysis of variance?

① $t$ distribution ② $F$ distribution
③ $\chi^2$ distribution ④ Normal distribution

9.5 Which is the correct process for the one-way ANOVA?

| |
|---|
| a. Calculate Total SS, Treatment SS, Error SS |
| b. Set the hypothesis |
| c. Test the hypothesis |
| d. Calculate the variance ration in the ANOVA table |
| e. Find the value in the F distribution table |

① a → b → c → d → e ② b → d → e → a → c
③ b → a → d → e → c ④ b → e → d → a → c

9.6 Which is the correct relationship between the total sum of squares (SST), between sum of squares (SSB), error sum of squares (SSE)?

① SST = SSB + SSE ② SST = SSB - SSE
③ SST = SSE - SSB ④ SST = SSB * SSE

9.7 If $F_{4,20;0.05} = 2.87$ and the observed $F$ ratio is 6.90 in the ANOVA table, what is your conclusion with the 5% significance level?

① significantly different ② no significant difference
③ very similar ④ unknown

9.8 Which is not appeared in the analysis of variance table?

① sum of squares ② F ratio
③ degrees of freedom ④ standard deviation

9.9 What is the name of variable which effects response variable in the experimental design?

① cause element ② independent variable
③ dependent variable ④ factor

9.10 In order to compare the fuel mileage of three types of cars, three drivers would like to drive cars, but fuel mileage may be affected by the driver. What is the name of variable like drivers?

① block variable ② independent variable
③ dependent variable ④ factor

9.11 When we compare the fuel mileage of three types of cars, which experimental desing is used to reduce the effect of drivers?

bvcxz① completely randomized design ② latin square method
③ two-way ANOVA ④ randomized block design

9.12 What is called the effect of a factor A that varies depending on the level of the factor B?

① main effect of factor A ② main effect of factor B
③ two-way ANOVA ④ interaction effect

(Answers)
9.1 ③, 9.2 ①, 9.3 ③, 9.4 ②, 9.5 ③, 9.6 ①, 9.7 ①, 9.8 ④, 9.9 ④, 9.10 ①,
9.11 ④, 9.12 ④