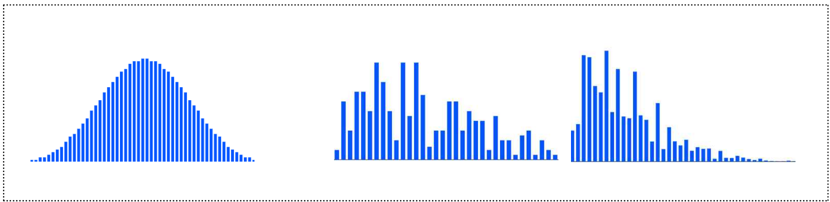
10

Nonparametric Testing Hypothesis



SECTIONS

CHAPTER OBJECTIVES

The hypothesis tests from Chapters 7 through 9 are based on assumptions such that the populations of continuous data follow the normal distributions. However, in real-world data, such assumptions may not be satisfied.

This chapter introduces the nonparametric methods for testing hypothesis by converting data such as rankings which do not require assumptions on the population distribution.

Section 10.1 introduces tests for the location parameter of single population such as the Sign Test and Signed Rank Test.

Section 10.2 introduces tests for comparing location parameters of two populations such as the Wilcoxon Rank Sum Test.

Section 10.3 introduces tests for comparing location parameter of several populations such as the Kruskal-Wallis Test and Friedman Test.

10.1 Nonparametric Test for the Location Parameter of Single Population

- The hypothesis test for a population mean in Chapter 7 can be done using t distribution in the case of a small sample if the population is assumed as a normal distribution. As such, if we make some assumptions about a population distribution and test a population parameter using sample data, it is called a parametric test. The hypothesis tests for two population parameters in Chapter 8 and the analysis variance in Chapter 9 are also parametric tests, because they assume that populations are normal distributions.
- However, real world data may not be appropriate to assume that a population follows a normal distribution, or there may not be enough number of samples to assume a normal distribution. In some cases, data collected are not continuous or can be ordinal such as rank, then the parametric tests are not appropriate. In such cases, methods to test population parameters by converting the data into signs or ranks without assuming on population distributions are called the distribution-free or nonparametric tests.
- Since the nonparametric test utilizes the converted data such as signs or ranks, there may be some loss of information about the data. Therefore, if a population can be assumed as a normal distribution, there is no reason to use the nonparametric tests. In fact, when a population follows a normal distribution, a nonparametric test has a higher probability of the type 2 error at the same significance level. However, a nonparametric test would be more appropriate if the data are from a population that do not follow a normal distribution.
- The hypothesis test for a population mean in Chapter 7 is based on the theory of the central limit theorem for the sampling distribution of all possible sample means. However, the nonparametric test use signs by examining whether data values are small or large from the central location parameter of the population (the Sign Test of 10.1.1), or use ranks by calculating the ranking of the data (the Wilcoxon Signed Rank Test of Section 10.1.2). Here, the central location parameter can be the population mean or the population median, but usually referring to the population median that is not affected by an extreme point of the data.
- Estimation of a population parameter can also be made by using a nonparametric method, but this chapter only introduces nonparametric hypothesis tests. Those interested in the nonparametric estimation should refer to the relevant literature.

10.1.1 Sign Test

- Let's take a look at the sign test with the following examples.

Example 10.1.1	A bag of cookies is marked with a weight of 200g. Ten bags are randomly selected from several retailers and examined their weights as follows. Can you say that there are as many cookies in the bag as the weight marked? 203 204 197 195 201 205 198 199 194 207 Ex ⇨ eBook ⇨ EX100101_CookieWeight.csv 1) Draw a histogram of the data to check whether a testing hypothesis using a parametric method can be performed. 2) Test the hypothesis by using a nonparametric method which utilizes the sign data by examining whether data values are smaller or larger than 200 with the significance level of 5%. 3) Check the result of the above test using 『eStatU』.

Example 10.1.1 Answer

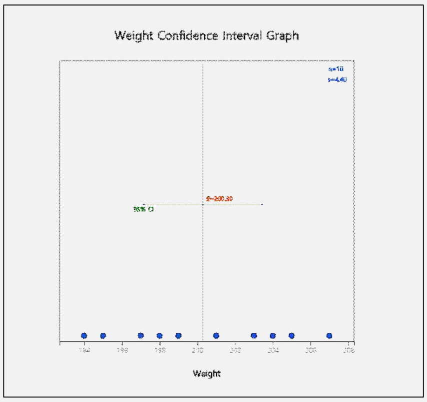
1) The null and alternative hypothesis to test the population mean μ can be written as follows:

$$H_0:\ \mu = 200, \quad H_1:\ \mu \neq 200$$

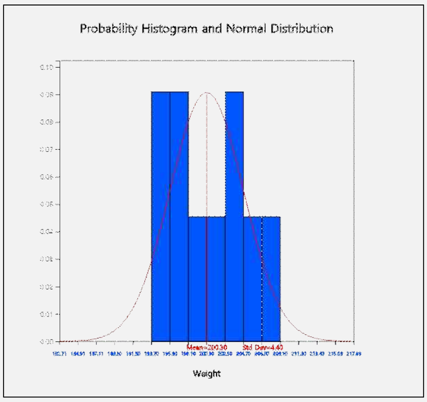
In order to test the hypothesis using the parametric t-test in Chapter 7, it is necessary to assume that the population is normally distributed, because the sample size of 10 is small. Let us check whether the sample data is a normal distribution by using a histogram. Enter data in 『eStat』 as shown in <Figure 10.1.1>

<Figure 10.1.1> Data input for cookie weight

- Click icon μ of the testing hypothesis for the population mean and select 'Weight' as the analysis variable in the variable selection box. A dot graph with the 95% confidence interval will appear as <Figure 10.1.2>. If you click the [Histogram] button in the options window below the graph, a histogram as shown in <Figure 10.1.3> will appear. If you look at the histogram, it is not sufficient to assume that the population follows a normal distribution. In such cases, applying the parametric hypothesis test may lead to errors.



<Figure 10.1.2> Dot graph of the cookie weight



<Figure 10.1.3> Histogram of the cookie weight

2) In this case, the sample data can be converted to sign data only by examining whether the weight of cookie bag is greater than 200g (+ marked) or not (- marked).

sample data	203	204	197	195	201	205	198	199	194	207
sign data	+	+	-	-	+	+	-	-	-	+

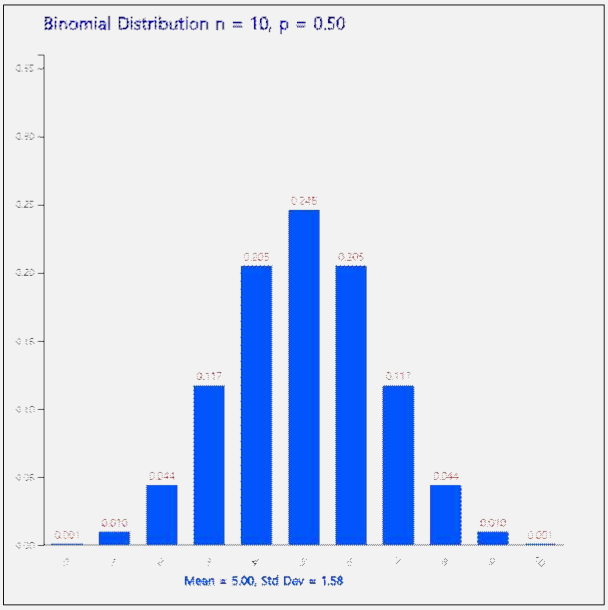
Example 10.1.1 Answer (continued)

If the number of + signs and − signs are similar, the weight of cookie bag would be 200g approximately. If the number of + signs is larger than − signs, then the weight of cookie bag is greater than 200g. If the number of − signs is larger than + signs, then the weight of cookie bag is less than 200g.

- Since the above sign data only investigate whether a data is larger or smaller than 200 and never use a concept of the mean, it can be considered as testing for the population median (M) as follows:

H_0: M = 200 $\quad\quad\quad$ H_1: $M \neq$ 200

- In the sign data above, 'the number of + signs' (denote it as n_+) or 'the number of − signs' (denote as n_-) follows a binomial distribution with parameters of n=10, p=0.5 (<Figure 10.1.4>).



<Figure 10.1.4> Binomial distribution when n=10, p=0.5

- Therefore, if H_0 is correct, the number of + signs may be the most likely to be 5 and 0, 1 or 9, 10 are very unlikely to be present. In order to test H_0: M = 200 with 5% significance level, since it is a two-sided test, rejection region should have the 2.5% probability at both ends of the binomial distribution, so it is approximately as follows:

  If the number of + signs (n_+) is either 0, 1 (cumulated probability from left is 0.011) or 9, 10 (cumulated probability from right is 0.011), then reject H_0

  This rejection region has a total probability of 2*0.011 = 0.022 which is smaller than the significance level of 0.05. When we use a discrete distribution such as binomial distribution, it may be difficult to find a rejection region which is exactly the same as the significance level. If we include one more value in the rejection region, the decision rule is as follows:

  If the number of + signs (n_+) is either 0, 1, 2 (cumulated probability from left is 0.055) or 8, 9, 10 (cumulated probability from right is 0.055), then reject H_0

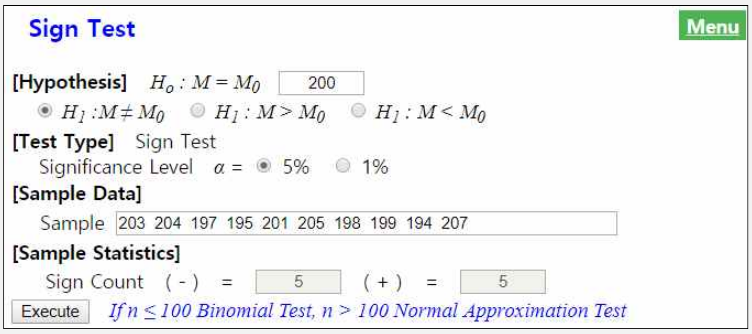
  This rejection region has a total probability of 2*0.055 = 0.110 which is greater than the significance leve of 0.05. Therefore, the middle values 1.5 (of 1 and 2) and 8.5 (of 8 and 9) can be used in the decision rule as follows:

  If the number of + signs (n_+) < 1.5 or n_+ > 8.5, then reject H_0

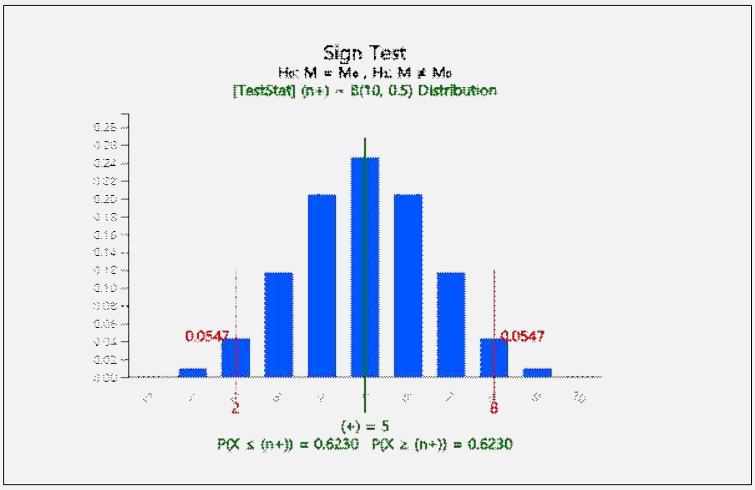
Example 10.1.1 Answer (continued)

This method may also be approximate. In the case of testing using a discrete distribution, it is not possible to say 'what is right' among the above decision rules and the analyst should select the critical value near the significance level. In this example, the number of + signs (n_+) is 5 and you can not reject the null hypothesis H_0. In other words, the median of the weight of the cookie bag is 200g.

3) Enter data as shown in <Figure 10.1.5> in 『eStatU』 and press the [Execute] button to show the test result as in <Figure 10.1.6>. It shows the critical lines for values containing the significance level of 5% (2.5% for both tests). For a discrete distribution such as the binomial distribution, the choice of the final reject region shall be determined by the analyst.



<Figure 10.1.5> Data input for sign test in 『eStatU』



<Figure 10.1.6> Result of sign test using 『eStatU』

[Practice 10.1.1]

A psychologist has selected 9 handicap workers randomly from production workers employed at various factories in a large industrial complex and their work competency scores are examined as follows. The psychologist wants to test whether the population median score is 40. Assume the population distribution is symmetrical about the mean.

32, 52, 21, 39, 23, 55, 36, 27, 37

Ex ⇨ eBook ⇨ PR100101_CompetencyScore.csv

1) Check whether a parametric test is possible.
2) Apply the sign test with the significance level of 5%.

- When the population median is M, the sign test is used to test whether $M = M_0$ or $M > M_0$ (or $M < M_0$ or $M \neq M_0$). However, if the population distribution is symmetrical to the mean, the sign test is the same as the test of the population mean, because mean and median are the same in this case.
- When there are n number of samples, the test statistic for the sign test uses the number of data which are greater than M_0 which is denoted as n_+. The sign test uses the random variable of 'the number of + signs (n_+)' which follows a binomial distribution with parameters n and p=0,5, i.e., $B(n,0.5)$ when the null hypothesis is true. You can use the number of data which are less than M_0, i.e., n_- = $n - n_+$ and n_- also follows a binomial distribution $B(n,0.5)$. Let us use n_+ in this section. $B(n,0.5)_\alpha$ represents the right tail 100$\times\alpha$ percentile, but the accurate percentile value may not exist, because it is a discrete distribution. In this case, middle value of two nearest percentile is often used. Table 10.1.1 summarizes the decision rule for each type of hypothesis of the sign test.

Table 10.1.1 Decision rule of the sign test

Type of Hypothesis	Decision Rule Test Statistic n_+ = 'number of plus sign data'
1) $H_0 : M = M_0$ $H_1 : M > M_0$	If $n_+ > B(n,0.5)_\alpha$, then reject H_0, else accept H_0
2) $H_0 : M = M_0$ $H_1 : M < M_0$	If $n_+ < B(n,0.5)_{1-\alpha}$, then reject H_0, else accept H_0
3) $H_0 : M = M_0$ $H_1 : M \neq M_0$	If $n_+ < B(n,0.5)_{1-\alpha/2}$ or $n_+ > B(n,0.5)_{\alpha/2}$, then reject H_0, else accept H_0

☞ **If the observed value is the same as M_0?**

If any of the observations has the same value as M_0, they are not used in the sign test. In other words, reduce n.

- As studied in Chapter 5, the binomial distribution $B(n,\,0.5)$ can be approximated by the normal distribution $N(0.5n,\,0.5^2 n)$ if n is sufficiently large. Therefore, if the sample size is large, the test statistic n_+ = 'the number of plus sign data' can be tested using the normal distribution $N(0.5n,\,0.5^2 n)$. Table 10.1.2 summarizes the decision rule for each hypothesis of the sign test in the case of large samples.

Table 10.1.2 Decision rule of the sign test (large sample case)

Type of Hypothesis	Decision Rule Test Statistic: n_+ = 'number of plus sign data'
1) $H_0 : M = M_0$ $H_1 : M > M_0$	If $\dfrac{n_+ - 0.5n}{\sqrt{0.25n}} > z_\alpha$, then reject H_0, else accept H_0
2) $H_0 : M = M_0$ $H_1 : M < M_0$	If $\dfrac{n_+ - 0.5n}{\sqrt{0.25n}} < -z_\alpha$, then reject H_0, else accept H_0
3) $H_0 : M = M_0$ $H_1 : M \neq M_0$	If $\left\lvert \dfrac{n_+ - 0.5n}{\sqrt{0.25n}} \right\rvert > z_{\alpha/2}$, then reject H_0, else accept H_0

10.1.2 Wilcoxon Signed Rank Sum Test

- The sign test described in the previous section converted sample data to either + or - symbols by examining whether the data were larger or smaller than the medium M_0. In this case, most of the information that the original sample data have is lost. In order to apply the Wilcoxon signed rank test, we subtract M_0 first from the sample data and take the absolute value of this data. Assign ranks on these absolute values and calculate the sum of the larger ranks than M_0 and the sum of the smaller ranks than M_0. If two rank sums are similar, we conclude that the population median is equal to M_0. This signed rank sum test is the most widely used nonparametric method for testing the central location parameter of a population. This test takes into account not only whether the sample data are larger or smaller than M_0, but also the relative size of the sample data from M_0.

Example 10.1.2

Using the cookie weight data of [Example 10.1.1], apply the signed rank test to see whether the weight of the cookie bag is 200g or not with the significance level of 5%

203 204 197 195 201 205 198 199 194 207
Ex ⇨ eBook ⇨ EX100101_CookieWeight.csv

Check the result of the signed rank test using 『eStatU』.

Example 10.1.2 Answer

- The hypothesis for this problem is to test whether the population median(M) is 200g or not.

$$H_0:\ M = 200 \qquad\qquad H_1:\ M \neq 200$$

- The signed rank sum test examines not only checking the sample data are greater than M_0 = 200g (+ sign) or not (- sign), but also checking the rank of values of |data − 200|. If there are tied values, assign the average rank to each of tied values. For example, since there are two tied values of '1' which is the smallest among |data − 200|, the corresponding ranks of 1 and 2 are averaged which is 1.5 and assign the averaged rank to each of value '1'.

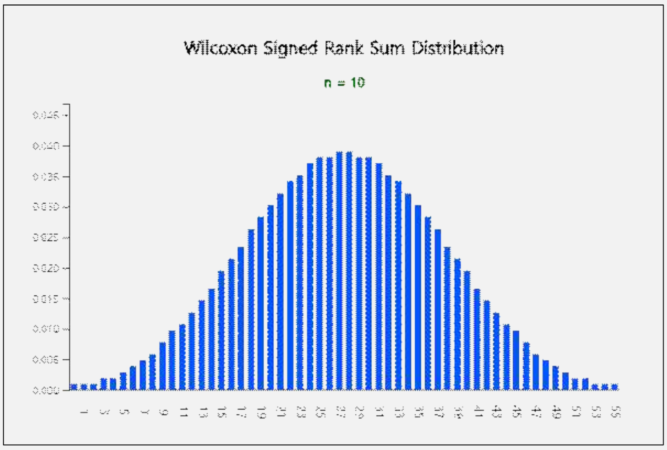
Sample data	203	204	197	195	201	205	198	199	194	207
Sign data	+	+	-	-	+	+	-	-	-	+
|data – 200|	3	4	3	5	1	5	2	1	6	7
Rank of |data – 200|	4.5	6	4.5	7.5	1.5	7.5	3	1.5	9	10
Rank sum of '+' sign (R_+)	4.5 + 6 + 1.5 + 7.5 + 10 = 29.5									

- The sum of all ranks is $1+2+\cdots+10 = 10(10+1)/2 = 55$. If the rank sum of + sign data (R_+) and the rank sum of − sign data (R_-) are similar (approximately 27.5 or so), the null hypothesis M = 200g would be true. In this example, R_+ = 29.5 and R_- = 25.5. Since R_+ is greater than R_-, the weight data which are greater than 200g appears to be dominant. What kind of large difference is statistically significant?
- To investigate how large a value is statistically significant when the null hypothesis is true, the sampling distribution of random variable R_+ = 'rank sum of + sign data' (or R_- = 'rank sum of − sign data') should be known. If H_0 is true, the number of cases for R_+ is shown in Table 10.1.3. It is not easy to examine all of these possible rankings to create a distribution table. 『eStatU』 shows the distribution of Wilcoxon signed rank sum as shown in <Figure 10.1.7> and its table as in Table 10.1.4.

Example 10.1.2 Answer (continued)

Table 10.1.3 All possible cases of R_+ = 'rank sum of + sign data'

Number of data with + sign	All possible combination of ranks	All possible rank sum of R_+
0	0	0
1	{1}, {2}, ⋯ , {10}	1, 2, ⋯ , 10
2	{1,2}, {1,3}, ⋯ , {1,10}, {2,3}, ⋯ , {2,10}, ⋯ {9,10}	3, 4, ⋯ , 11, 5, ⋯ , 12, ⋯ 19
⋯	⋯	⋯
10	{1,2, ⋯ ,10}	55



<Figure 10.1.7> Distribution of Wilcoxon signed rank sum when $n=10$

Table 10.1.4 Distribution of Wilcoxon signed rank sum when $n=10$

Wilcoxon Signed Rank Sum Distribution	n = 10		
x	P(X = x)	P(X ≤ x)	P(X ≥ x)
0	0.0010	0.0010	1.0000
1	0.0010	0.0020	0.9990
2	0.0010	0.0029	0.9980
3	0.0020	0.0049	0.9971
4	0.0020	0.0068	0.9951
5	0.0029	0.0098	0.9932
6	0.0039	0.0137	0.9902
7	0.0049	0.0186	0.9863
8	0.0059	0.0244	0.9814
9	0.0078	0.0322	0.9756
⋯	⋯	⋯	⋯
47	0.0059	0.9814	0.0244
48	0.0049	0.9863	0.0186
49	0.0039	0.9902	0.0137
50	0.0029	0.9932	0.0098
51	0.0020	0.9951	0.0068
52	0.0020	0.9971	0.0049
53	0.0010	0.9980	0.0029
54	0.0010	0.9990	0.0020
55	0.0010	1.0000	0.0010

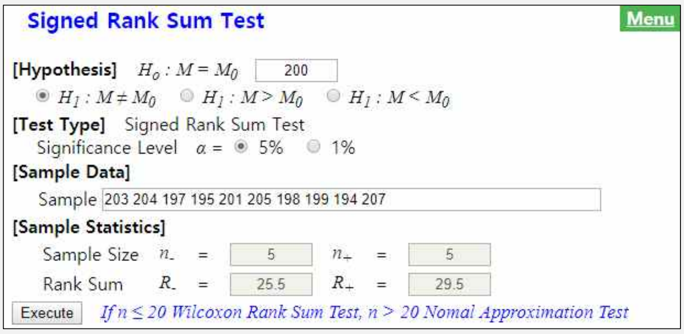
Example 10.1.2 Answer (continued)

- Since it is a two-sided test with the 5% significance level, if you find a 2.5% percentile at both ends, P(X $\leq$ 8) = 0.0244, P(X $\geq$ 47) = 0.0244. In case of a discrete distribution, we can not find the exact 2.5 percentile from both ends. Therefore, the decision rule can be written as follows:

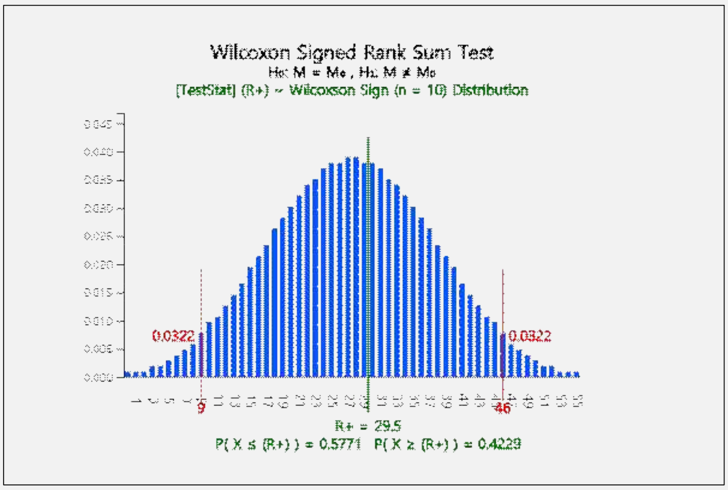
  'If $R_+ \leq 8.5$ or $R_+ \geq 46.5$, then reject H_0'

  Since R_+ = 29.5 in this problem, we can not reject H_0.

- After entering the data in 『eStatU』 as in <Figure 10.1.8>, pressing the [Execute] button will calculate the sample statistics and show the test result as in <Figure 10.1.9>. The critical lines are the value for containing 5% significance level from both sides (the probability of each end is 2.5%). For a discrete distribution, the choice of the final reject region should be determined by the analyst.



<Figure 10.1.8> 『eStatU』 Signed rank sum test

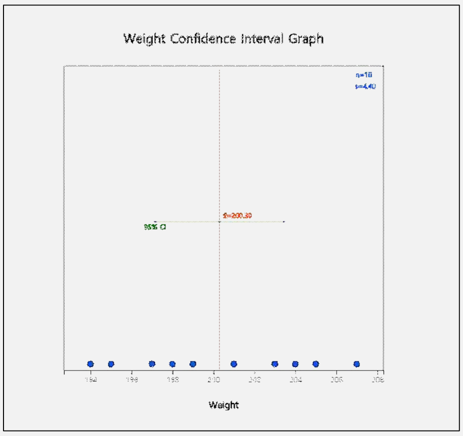


<Figure 10.1.9> Signed rank sum test in 『eStatU』

- The signed rank sum test may be done using 『eStat』. If you enter the data as shown in <Figure 10.1.10>, select 'Weight' as the analysis variable in the variable selection box and click the icon $\boxed{\mu}$ of testing the population mean. Then a dot graph with the 95% confidence interval for the population mean will appear as <Figure 10.1.11>.

Example 10.1.2 Answer (continued)

<Figure 10.1.10> Data input for cookie weight



<Figure 10.1.11> Dot graph and confidence interval of cookie weight

- Enter a value of 200 from the options below the graph and click the [Wilcoxon Signed Rank Sum Test] button to display the same test result graph and result table as in <Figure 10.1.12>.

Wilcoxon Signed Rank Sum Test	Analysis Var	Weight			
Statistics	Total Observation	Test Observation	-Group Observation	+Group Observation	
	10	10	5	5	
Missing Observations	0				
Hypothesis					
H_0 : M = M_0	μ_0	[TestStat]	Rank Sum R+	P(X ≤ R+)	P(X ≥ R+)
H_1 : M ≠ M_0	200.00	+Group Rank Sum(R+)	29.50	0.5771	0.4229

<Figure 10.1.12> Result of the Wilicoxon Signed Rank Sum Test

- If we denote the population median as M, the signed rank sum test is to test whether the population median is M_0 or greater than (or less than or not equal to) M_0. However, if the population distribution is symmetric about the mean, the signed rank sum test becomes to test about the population mean because the population median and mean are the same. The basic statistical model is as follows:

  $X_i = M_0 + \epsilon_i, \ i = 1,2,\cdots,n$
  where ϵ_i's are independent, symmetric about the mean 0 and follow the same distribution.

- If $x_1, x_2, \ldots, x_n$ are sample data, ranks of $|x_i - M_0|$ are calculated first and the sum of ranks for the data which are greater than M_0 (+ sign data of $x_i - M_0$), denoted as R_+, is calculated. R_+ is the test statistic for the signed rank sum test and the sampling distribution of R_+, denoted as $w_+(n)$, is calculated for testing hypothesis by considering all possible cases. 『eStatU』 provides $w_+(n)$ until $n=22$. $w_+(n)_\alpha$ denotes right tail 100×α percentile of the $w_+(n)$ distribution, but it is not easy to find the exact percentile because $w_+(n)$ is a discrete distribution and is usually used to approximate the two adjacent values. Table 10.1.5 summarizes the decision rule for the Wilcoxon signed rank sum test for each type of hypothesis.

Table 10.1.5 Decision rule of Wilcoxon signed rank sum test

Type of Hypothesis	Decision Rule Test Statistic R_+ = Rank sum of + sign data of $\mid x_i - M_0 \mid$
1) $H_0 : M = M_0$ $H_1 : M > M_0$	If $R_+ > w_+(n)_\alpha$, then reject H_0, else accept H_0
2) $H_0 : M = M_0$ $H_1 : M < M_0$	If $R_+ < w_+(n)_{1-\alpha}$, then reject H_0, else accept H_0
3) $H_0 : M = M_0$ $H_1 : M \neq M_0$	If $R_+ < w_+(n)_{1-\alpha/2}$ or $R_+ > w_+(n)_{\alpha/2}$, then reject H_0, else accept H_0

☞

If the observed value is the same as M_0?

If any of the observed values has the same value as M_0, they are not used in the test. In other words, reduce n.

[Practice 10.1.2]

A psychologist has selected 9 handicap workers randomly from production workers employed at various factories in a large industrial complex and their work competency scores are examined as follows. The psychologist wants to test whether the population median score is 45. Assume the population distribution is symmetrical about the mean.

32, 52, 21, 39, 23, 55, 36, 27, 37

Ex ⇨ eBook ⇨ PR100101_CompetencyScore.csv

1) Check whether a parametric test is possible.
2) Apply the Wilcoxon signed rank test with the significance level of 5%.
3) Compare this test result with the sign test of [Practice 10.1.1].

- If the sample size is large enough, the test statistic R_+ is approximated to a normal distribution with the following mean $E(R_+)$ and variance $V(R_+)$ when the null hypothesis is true.

$$E(R_+) = \frac{n(n+1)}{4}$$

$$V(R_+) = \frac{n(n+1)(2n+1)}{24}$$

- Table 10.1.6 summarizes the decision rule of the signed rank sum test for each type of hypothesis.

Table 10.1.6 Decision rule of Wilcoxon signed rank sum test (large sample case)

Type of Hypothesis	Decision Rule Test Statistic: R_+ = Rank sum of + sign data of $\lvert x_i - M_0 \rvert$
1) $H_0 : M = M_0$ $H_1 : M > M_0$	If $\dfrac{R_+ - E(R_+)}{\sqrt{V(R_+)}} > z_\alpha$, then reject H_0, else accept H_0
2) $H_0 : M = M_0$ $H_1 : M < M_0$	If $\dfrac{R_+ - E(R_+)}{\sqrt{V(R_+)}} < -z_\alpha$, then reject H_0, else accept H_0
3) $H_0 : M = M_0$ $H_1 : M \neq M_0$	If $\left\lvert \dfrac{R_+ - E(R_+)}{\sqrt{V(R_+)}} \right\rvert > z_{\alpha/2}$, then reject H_0, else accept H_0

- The distribution of $w_+(n)$ is independent of the population distribution. In other words, the Wilcoxon signed rank sum test is a distribution free test. For example, if $n=3$, the distribution of $w_+(3)$ can be obtained as follows:

Rank			Possible value of R_+
1	2	3	
-	-	-	0
+	-	-	1
-	+	-	2
-	-	+	3
+	+	-	3
+	-	+	4
-	+	+	5
+	+	+	6

- Therefore, the distribution of $w_+(3)$ can be calculated as follows which is independent of the population distribution.

$R_+ = x$	0	1	2	3	4	5	6
$P(R_+ = x)$	$\frac{1}{8}$	$\frac{1}{8}$	$\frac{1}{8}$	$\frac{2}{8}$	$\frac{1}{8}$	$\frac{1}{8}$	$\frac{1}{8}$

- If there is a tie on the value of $\lvert x_i - M_0 \rvert$, the average rank is calculated when the ranking is obtained. In this case, the variance of R_+ in case of large sample is calculated using the following modified formula.

$$V(R_+) = \frac{1}{24}\left[n(n+1)(2n+1) - \frac{1}{2}\sum_{j=1}^{g} t_j(t_j-1)(t_j+1)\right]$$

Here g = (number of groups of tie)

t_j = (size of j^{th} tie group, i.e., number of observations in the tie group)

if there is no tie, size of j^{th} tie group is 1 and t_j=1

10.2 Nonparametric Test for Location Parameters of Two Populations

- The testing hypothesis for the two population means in Chapter 8 used the t-distribution in case of a small sample, if each population could be assumed to be a normal distribution. However, the assumption that the population follows a normal distribution may not be appropriate for real world data, or that there may not be enough sample data to assume a normal distribution. Alternatively, if collected data is ordinal such as ranking, then the parametric t-test is not appropriate. In such cases, a nonparametric method is used to test parameters by converting data to ranks without assuming the distribution of the population. This section introduces the Wilcoxon rank sum test.
- Nonparametric tests convert data into ranks, so there may be some loss of information about the data. Therefore, if data are normally distributed, there is no reason to apply a nonparametric test. However, a nonparametric method would be a more appropriate method if the data do not follow a normal distribution.
- As in Chapter 8, this section introduces nonparametric tests for testing location parameters of two populations for the samples drawn independently from each population and for the samples drawn as paired.

10.2.1 Independent Samples: Wilcoxon Rank Sum Test

- Let's take a look at the Wilcoxon rank sum test with the following example.

Example 10.2.1

A professor teaches Statistics courses to students in the Department of Economics and the Department of Management. In order to compare exam scores of students in the two departments, seven students were randomly sampled from the Economics Department and six students from the Management Department and their scores were as follows:

Department of Economics 87 75 65 95 90 81 93
Department of Management 57 85 90 83 87 71
⇨ eBook ⇨ EX100201_ScoreByDepartment.csv

1) Draw a histogram of the data to verify that the testing hypothesis can be performed using a parametric method.
2) Apply the Wilcoxon rank sum test with the significance level of 5%.
3) Check the result of the Wilcoxon rank sum test using 『eStat』.

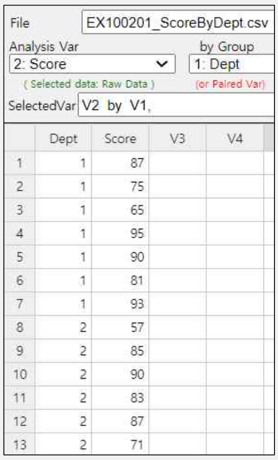
Example 10.2.1 Answer

1) The hypothesis of this problem to test two population means μ_1 and μ_2 are as follows:

$$H_0 : \mu_1 = \mu_2, \qquad H_1 : \mu_1 \neq \mu_2$$

Since the sample sizes, n_1 = 7 and n_2 = 6, are small from each population, it is necessary to assume that the populations are normally distributed in order to apply the parametric t-test. In order to check whether each sample data follows a normal distribution, let us draw a histogram using 『eStat』. Enter data in 『eStat』 as shown in <Figure 10.2.1>.

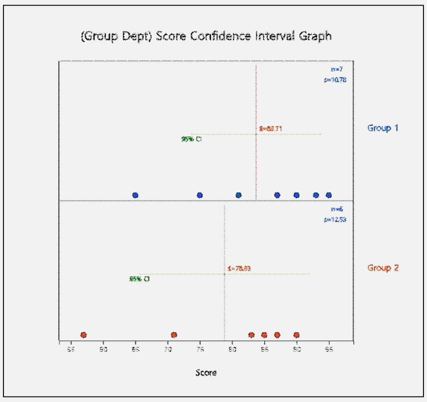
Example 10.2.1 Answer (continued)


File: EX100201_ScoreByDept.csv
Analysis Var: 2: Score | by Group: 1: Dept
(Selected data: Raw Data) (or Paired Var)
SelectedVar: V2 by V1,

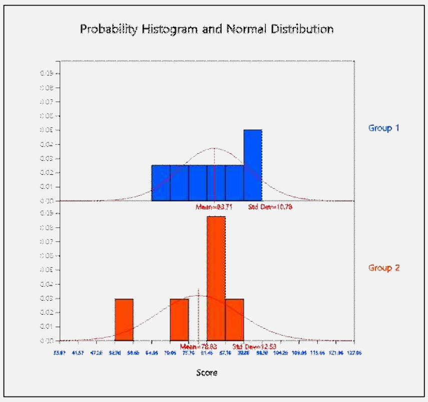
	Dept	Score	V3	V4
1	1	87		
2	1	75		
3	1	65		
4	1	95		
5	1	90		
6	1	81		
7	1	93		
8	2	57		
9	2	85		
10	2	90		
11	2	83		
12	2	87		
13	2	71		

<Figure 10.2.1> Data input at 『eStat』

- Click icon for testing two population means in the main menu. Select ‘Score’ as 'Analysis Var' and ‘Dept’ as ‘By Group’ variable. Then, two dot graphs together with 95% confidence intervals for each population mean will appear as in <Figure 10.2.2>. Average score of students in the Economics Department appears to be higher than the average score of students in the Management Department, but it should be tested for statistical significance. Pressing the [Histogram] button in the options window below the graph will reveal the histogram and normal distribution curves for each department as in<Figure 10.2.3>.



<Figure 10.2.2> Dot graph and confidence interval by department



<Figure 10.2.3> Histogram by department

2) Looking at the histogram, the small number of data is not sufficient to assume that the population follows a normal distribution. In such case, applying the parametric t-test may lead to error. In case of a nonparametric test, we test the location parameter of the population such as median which is not so sensitive to extreme values. The hypothesis for this problem is to test whether the median values M_1 and M_2 of the two populations are equal or not as follows:

$$H_0 : M_1 = M_2 \qquad H_1 : M_1 \neq M_2$$

- The Wilcoxon rank sum test calculates ranks of each data by combining two samples first and then calculate the sum of ranks in each sample. If there is a tie, then the averaged rank shall be used. To obtain the ranks of the combined sample, it is convenient to arrange each sample data in ascending order as shown in Table 10.2.1. The sum of ranks R_1 and R_2 in each sample will be used as the test statistic for the Wilcoxon rank sum test.

Example 10.2.1 Answer (continued)

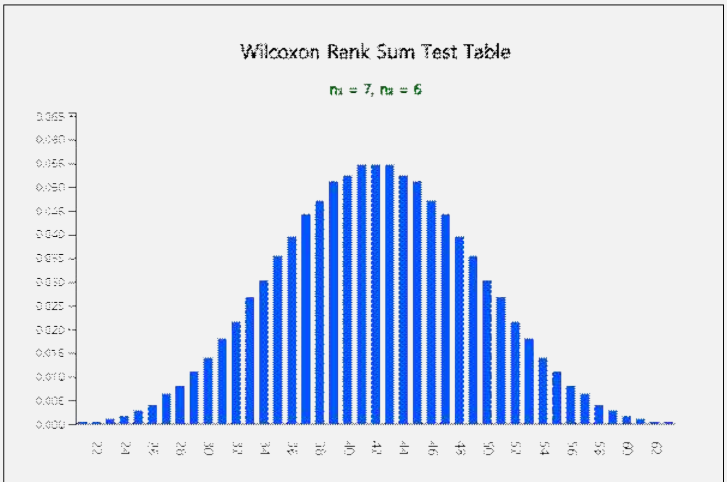
Table 10.2.1 A table to calculate ranks in a combined sample

Sorted Data of Sample 1	Sorted Data of Sample 2	Ranks of Sample 1	Ranks of Sample 2
	57		1
65		2	
	71		3
75		4	
81		5	
	83		6
	85		7
87	87	8.5	8.5
90	90	10.5	10.5
93		12	
95		13	
	Sum of ranks	R_1 = 55	R_2 = 36

- The sum of all ranks is $1+2+\cdots+13 = 13(13+1)/2 = 91$. The sum of ranks in sample 1 is R_1 = 55 and the sum of ranks in sample 2 is R_2 = 36. Note that R_1 + R_2 = 91. If R_1 and R_2 are similar, the null hypothesis that two population medians are the same is accepted. In this example R_1 is larger than R_2 and it seems the median of the population 1 is larger than the median of the population 2. But how much difference in the rank sum would be statistically significant if you consider the sample sizes?
- To investigate how large a difference in the rank sum is statistically significant when the null hypothesis is true, the sampling distribution of the random variable R_2 = 'Rank sum of sample 2' (or R_1 = 'Rank sum of sample 1') should be known. If H_0 is true, the number of cases for R_2 is ${}_{13}P_6 = 1716$ as shown in Table 10.2.2. It is not easy to examine all of these possible rankings to find the distribution table. 『eStatU』 provides the Wilcoxon rank sum distribution and its table as shown in <Figure 10.2.4>.

Table 10.2.2 All possible ranks for six data in sample 2 if n = 13

All possible permutation of ranks	Sum of ranks, R_2
{1,2,3,4,5,6}	21
{1,2,3,4,5,7}	22
…	…
{8,9,10,11,12,13}	63



<Figure 10.2.4> Wilcoxon rank sum distribution when n_1=7, n_2=6

Example 10.2.1 Answer (continued)

Table 10.2.3 Wilcoxon rank sum distribution when n_1=7, n_2=6

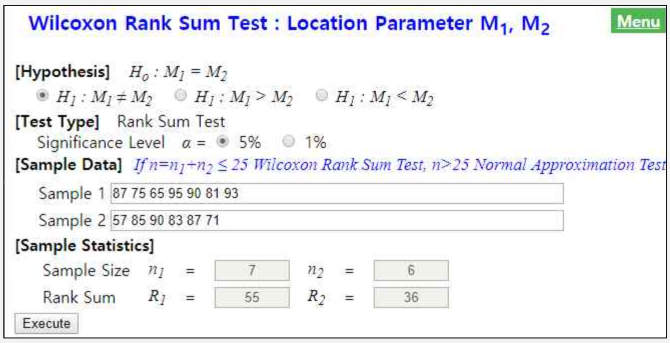
Wilcoxon rank sum distribution	n_1 = 7	n_2 = 6	
x	P(X = x)	P(X $\leq$ x)	P(X $\geq$ x)
21	0.0006	0.0006	1
22	0.0006	0.0012	0.9994
23	0.0012	0.0023	0.9988
24	0.0017	0.0041	0.9977
25	0.0029	0.007	0.9959
26	0.0041	0.0111	0.993
27	0.0064	0.0175	0.9889
28	0.0082	0.0256	0.9825
29	0.0111	0.0367	0.9744
…	…	…	…
55	0.0111	0.9744	0.0367
56	0.0082	0.9825	0.0256
57	0.0064	0.9889	0.0175
58	0.0041	0.993	0.0111
59	0.0029	0.9959	0.007
60	0.0017	0.9977	0.0041
61	0.0012	0.9988	0.0023
62	0.0006	0.9994	0.0012
63	0.0006	1	0.0006

- Since the hypothesis requires a two sided test with the significance level of 5%, so if you find a 2.5 percentile at both ends, P(X $\leq$ 28) = 0.0256, P(X $\geq$ 56) = 0.0256. Since it is a discrete distribution, there is no exact value of the 2.5 percentile. Therefore, the decision rule can be set as follows:

  'If $R_2 \leq 27.5$ or $R_2 \geq 56.5$, then reject H_0'

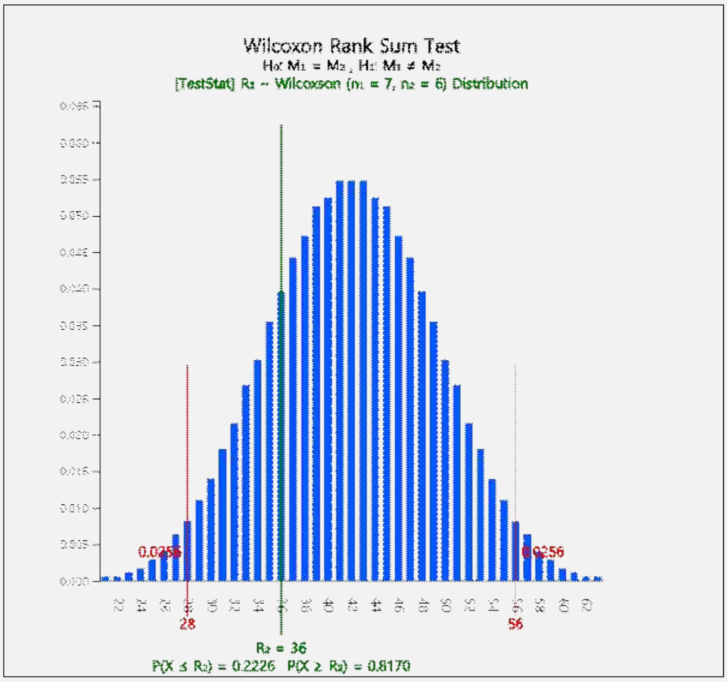
  In this problem R_2 = 36, and therefore, we cannot reject H_0 which means the difference between R_1 and R_2 is not statistically significant.

3) In『eStatU』, enter the data as <Figure 10.2.5> and click the [Execute] button. It will calculate the sample statistics and show the test result graph as <Figure 10.2.6>. The two critical lines which correspond to 2.5% from the end are shown here. For a discrete distribution such as this, the choice of the final rejection region should be determined by the analyst.



<Figure 10.2.5> Data input for the Wilcoxon rank sum test at『eStatU』

Example 10.2.1 Answer (continued)



<Figure 10.2.6> Wilcoxon rank sum test using 『eStatU』

- The rank sum test can be performed using 『eStat』. After you saw <Figure 10.2.2>, click the [Wilcoxon Rank Sum Test] button in the options window below the graph. Then a test result graph as shown in <Figure 10.2.6> will appear in the Graph Area and a test result table as in <Figure 10.2.7> will appear in the Log Area.

Wilcoxon Rank Sum Test	Analysis Var	Score			
Statistics	Observation	Mean	Std Dev	Rank Sum	
1 (Group 1)	7	83.714	10.781	55.00	
2 (Group 2)	6	78.833	12.529	36.00	
Total	13	81.462	11.399	91.00	
Missing Observations	0				
Hypothesis					
$H_0: M_1 - M_2 = D$	D	[TestStat]	Group 2 Rank Sum R_2	$P(X \le R_2)$	$P(X \ge R_2)$
$H_1: M_1 - M_2 \ne D$	0.00	Group 2 Rank Sum(R_2)	36.00	0.2226	0.8170

<Figure 10.2.7> Result table of Wilcoxon rank sum test

- Let's generalize the Wilcoxon rank sum test described in [Example 10.2.1]. Denote random samples selected independently from each of the two populations as follows. The sample sizes are n_1 and n_2 respectively, and $n = n_1 + n_2$.

  Sample 1 $\quad X_1, X_2, \cdots, X_{n_1}$

  Sample 2 $\quad Y_1, Y_2, \cdots, Y_{n_2}$

  For convenience, assume $n_1 \ge n_2$. If $n_1 \le n_2$, you can swap between X and Y.

- The statistical model of the Wilcoxon rank sum test is as follows:

  $X_i = M_1 + \epsilon_i, \quad i = 1,2,\cdots,n_1$
  $Y_j = M_1 + \Delta + \epsilon_j, \; j = 1,2,\cdots,n_2,$ * You may write $M_2 = M_1 + \Delta$

  Here Δ is the difference between location parameters. ϵ_i's are independent and follow the same continuous distribution which is symmetric around 0.
- The test statistic for the Wilcoxon rank sum test is the sum of ranks, R_2, for $Y_1, Y_2, \cdots, Y_{n_2}$ based on the combined sample of $X_1, X_2, \cdots, X_{n_1}, Y_1, Y_2, \cdots, Y_{n_2}$. The distribution of the random variable R_2 = 'Sum of the ranks for Y sample' can be obtained by investigating all possible cases of ranks for Y which is ${}_nP_{n_2}$ and is denoted as $w_2(n_1,n_2)$. 『eStatU』 provides the Wilcoxon rank sum distribution $w_2(n_1,n_2)$ and its table up to $n = 25$. $w_2(n_1,n_2)_\alpha$ denotes the right tail $100 \times \alpha$ percentile, but it might not be able to find the accurate percentile, because $w_2(n_1,n_2)$ is a discrete distribution. In this case, middle value of two percentiles near $w_2(n_1,n_2)_\alpha$ is often used as an approximation. Table 10.2.4 summarizes the decision rule for each type of hypothesis.

Table 10.2.4 Wilcoxon rank sum test

Type of Hypothesis	Decision Rule Test Statistic: R_2 = 'Sum of ranks assigned samples of Y'
1) $H_0 : M_1 = M_2$ $H_1 : M_1 > M_2$	If $R_2 > w_2(n_1,n_2)_\alpha$, then reject H_0, else accept H_0
2) $H_0 : M_1 = M_2$ $H_1 : M_1 < M_2$	If $R_2 < w_2(n_1,n_2)_{1-\alpha}$, then reject H_0, else accept H_0
3) $H_0 : M_1 = M_2$ $H_1 : M_1 \neq M_2$	If $R_2 < w_2(n_1,n_2)_{1-\alpha/2}$ or $R_2 > w_2(n_1,n_2)_{\alpha/2}$, then reject H_0, else accept H_0

☞ If there is a tie in the combined sample, assign the average rank.

[Practice 10.2.1]

A company wants to compare two methods of obtaining information about a new product. Among company employees, 17 employees were randomly selected and divided into two groups. The first group learned about the new product by the method A, and the second group learned by the method B. At the end of the experiment, the employees took a test to measure their knowledge of the new product and their test scores are as follows:

Method A:	50	59	60	71	80	78	72	77	73
Method B:	52	54	58	78	65	61	60	72	

Ex ⇨ eBook ⇨ PR100201_ScoreByMethod.csv

1) Can we apply a parametric test to conclude that population means of the two groups are different?
2) Apply a nonparametric test to conclude that the median values of the two groups are different. Test with the significance level of 0.05.

- When the null hypothesis is true, if the sample is large enough, the test statistic is approximated to the normal distribution with the following mean $E(R_2)$ and variance $V(R_2)$:

$$E(R_2) = \frac{n_2(n_1+n_2+1)}{2}$$
$$V(R_2) = \frac{n_1 n_2(n_1+n_2+1)}{12}$$

- Table 10.2.5 summarizes the decision rule for each hypothesis type of the Wilcoxon rank sum test if the sample is large enough.

Table 10.2.5 Wilcoxon rank sum test (large sample case)

Type of Hypothesis	Decision Rule Test Statistic: R_2 = 'Sum of ranks assigned samples of Y'
1) $H_0 : M_1 = M_2$ $H_1 : M_1 > M_2$	If $\frac{R_2 - E(R_2)}{\sqrt{V(R_2)}} > z_\alpha$, then reject H_0, else accept H_0
2) $H_0 : M_1 = M_2$ $H_1 : M_1 < M_2$	If $\frac{R_2 - E(R_2)}{\sqrt{V(R_2)}} < -z_\alpha$, then reject H_0, else accept H_0
3) $H_0 : M_1 = M_2$ $H_1 : M_1 \neq M_2$	If $\left\lvert \frac{R_2 - E(R_2)}{\sqrt{V(R_2)}} \right\rvert > z_{\alpha/2}$, then reject H_0, else accept H_0

- The distribution of rank sum statistic, $w_2(n_1,n_2)$, is not dependent on the population distribution. That is, the rank sum test is a distribution free test. For example, if n_1 = 3 and n_2 = 2, the distribution $w_2(3,2)$ can be found as follows. All possible cases of ranks for R_2 is ${}_5P_2 = 10$.

All possible ranks for combined sample					Value of R_2
X_1	X_2	X_3	Y_1	Y_2	
3	4	5	1	2	3
2	4	5	1	3	4
2	3	5	1	4	5
2	3	4	1	5	6
1	4	5	2	3	5
1	3	5	2	4	6
1	3	4	2	5	7
1	2	5	3	4	7
1	2	4	3	5	8
1	2	3	4	5	9

- Therefore, the distribution $w_2(3,2)$ is given regardless of the population distribution as follows:

$R_2 = x$	3	4	5	6	7	8	9
$P(R_2 = x)$	$\frac{1}{10}$	$\frac{1}{10}$	$\frac{2}{10}$	$\frac{2}{10}$	$\frac{2}{10}$	$\frac{1}{10}$	$\frac{1}{10}$

- If there is a tie in the combined sample, the average rank is assigned to each data. In this case, the variance of R_2 should be modified in case of large sample as follows:

$$V(R_2) = \frac{n_1 n_2}{12}\left[n_1+n_2+1-\frac{\sum_{j=1}^{g} t_j(t_j-1)(t_j+1)}{(n_1+n_2)(n_1+n_2-1)}\right]$$

Here g = (number of tied groups)
t_j = (size of j^{th} tie group, i.e., number of observations in the tie group)
if there is no tie, size of j^{th} tie group is 1 and t_j=1

10.2.2 Paired Samples: Wilcoxon Signed Rank Sum Test

- Section 8.1.2 discussed the testing hypothesis for two population means using paired samples. Paired samples are used when it is difficult to extract samples independently from two populations, or if independently extracted, the characteristics of each sample object are so different that the resulting analysis is meaningless. If two populations are normally distributed, the t-test was applied for the difference data of the paired samples as described in Section 8.1.2. However, if the normality assumption of two populations can not be satisfied, the Wilcoxon signed rank sum test in Section 10.1.2, which is a nonparametric test, can be applied to the difference data of the paired samples.
- In case of the paired samples, first calculate the differences ($d_i = x_{i1} - x_{i2}$) for each paired sample as shown in Table 10.2.6. For the data of differences, we examine the normality to check whether the parametric test can be applicable or not. If it is not applicable, we apply the Wilcoxon signed rank sum test on the differences.

Table 10.2.6 Data of differences for paired samples

Pair number	Sample of population 1 x_{i1}	Sample of population 2 x_{i2}	Difference $d_i = x_{i1} - x_{i2}$
1	x_{11}	x_{12}	$d_1 = x_{11} - x_{12}$
2	x_{21}	x_{22}	$d_2 = x_{21} - x_{22}$
...	...	...	...
n	x_{n1}	x_{n2}	$d_n = x_{n1} - x_{n2}$

- Let's take a look at the next example.

Example 10.2.2	The following is the survey result of eight samples from young couples. The husband's age and wife's age of each couple are recorded. (28, 28) (30, 29) (34, 31) (29, 32) (28, 29) (31, 33) (39, 35) (34, 29) Ex ⇨ eBook ⇨ EX100202_AgeOfCouple.csv 1) Calculate data of differences in each pair and draw their histogram to check whether a parametric test is applicable or not. 2) Apply the Wilcoxon signed rank sum test to see whether the husband's age is greater than the wife's age with the significance level of 0.05. 3) Check the result of the above signed rank sum test using 『eStat』.

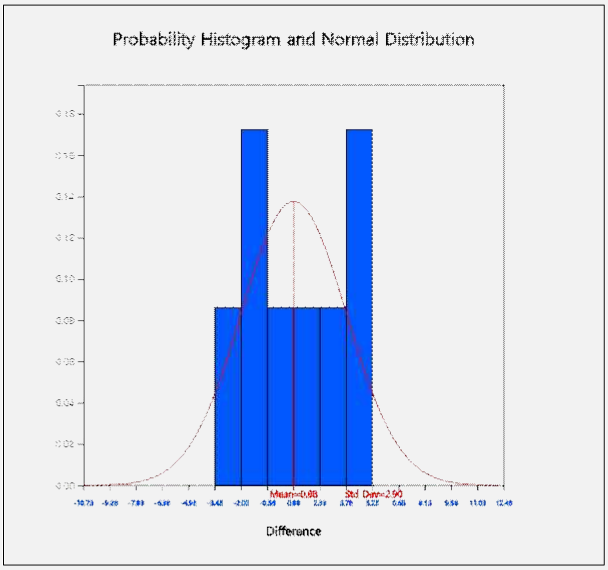
Example 10.2.2 Answer

1) The data of age differences between husband and wife are as follows:

Table 10.2.7 Data of age differences between husband and wife

Number	Husband x_i	Wife y_i	Difference $d_i = x_i - y_i$
1	28	28	0
2	30	29	1
3	34	31	3
4	29	32	-3
5	28	29	-1
6	31	33	-2
7	39	35	4
8	34	29	5

- The histogram for the data of differences by using 『eStat』 μ (the testing hypothesis for a population mean) is as in <Figure 10.2.8>. If you look at the histogram, it is not sufficient evidence that the data of differences follow a normal distribution, because the number of data is small. In such a case, applying the parametric hypothesis test may lead to errors. An appropriate nonparametric method for this problem is the Wilcoxon signed rank sum test on the data of differences.



<Figure 10.2.8> Histogram of age difference

2) The hypothesis to test is that the population median of the husband's age (M_1) is the same as the population median of the wife's age (M_2) or not as follows:

$$H_0: M_1 = M_2 \qquad H_1: M_1 \neq M_2$$

Since it is a paired sample, the hypothesis can be written whether the population median of differences (M_d) is equal to 0 or not as follows:

$$H_0: M_d = 0 \qquad H_1: M_d \neq 0$$

- In order to apply the signed rank sum test on the data of differences, we count the number of differences which is greater than 0 (denote as + sign) or not (denote as − sign) and assign ranks on |difference − 0|. Then calculate the sum of ranks with + sign and the sum of ranks with − sign. If the difference data is 0, omit the data. If there are ties on the difference data, assign the average rank.

Example 10.2.2 Answer (continued)

Difference data	1	3	-3	-1	-2	4	5
Sign data	+	+	-	-	-	+	+
|data – 0|	1	3	3	1	2	4	5
Rank of |data – 0|	1.5	4.5	4.5	1.5	3	6	7
Rank sum of '+' sign (R_+= 19)	1.5	+ 4.5				+ 6	+ 7

- In 『eStatU』, the distribution of the Wilcoxon signed rank sum when $n=7$ is shown in <Figure 10.2.9> and Table 10.2.8.

Table 10.2.8 Wilcoxon signed rank sum distribution when $n=7$

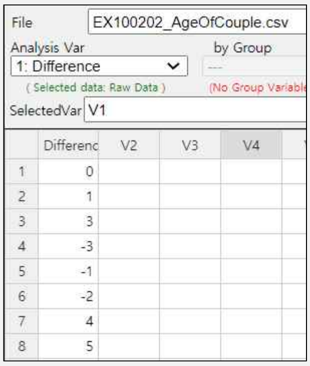
Wilcoxon Signed Rank Sum Distribution	n = 7		
x	P(X = x)	P(X $\le$ x)	P(X $\ge$ x)
0	0.0078	0.0078	1.0000
1	0.0078	0.0156	0.9922
2	0.0078	0.0234	0.9844
3	0.0156	0.0391	0.9766
…	…	…	…
25	0.0156	0.9766	0.0391
26	0.0078	0.9844	0.0234
27	0.0078	0.9922	0.0156
28	0.0078	1.0000	0.0078

- Since it is a two-sided test with the significance level of 5%, if a 2.5 percentile is found at both ends, P(X $\le$ 2) = 0.0234, P(X $\ge$ 26) = 0.0234. Since it is a discrete distribution, there is no exact value of the 2.5 percentile. Therefore, the decision rule is as follows:

  If $R_+ \le 2.5$ or $R_+ \ge 25.5$, reject H_0

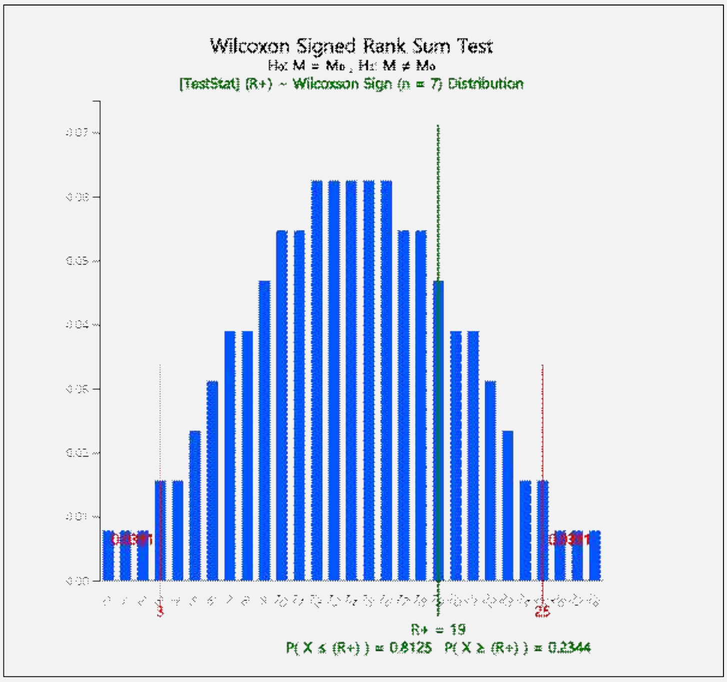
  Since R_+ = 19 in this problem, we can not reject the null hypothesis H_0 and conclude that the husband's age and the wife's age are the same.

3) Enter the data as shown in <Figure 10.2.9> in 『eStat』 and click the icon μ which is the test for a population mean. If you select the variable 'Difference' as the analysis variable, a dot graph with the 95% confidence interval for the population mean difference will appear. If you enter 0 for testing value in the hypothesis option and click the [Execute] button, you will see the test result as in <Figure 10.2.10> and <Figure 10.2.11>. Two critical lines for values containing 2.5 percentile from both sides are shown here. For a discrete distribution, the choice of the final decision rule should be determined by the analyst.



<Figure 10.2.9> Data difference

Example 10.2.2 Answer (continued)



<Figure 10.2.10> 『eStat』 Signed rank sum test

Wilcoxon Signed Rank Sum Test	Analysis Var	Difference			
Statistics	Total Observation	Test Observation	-Group Observation	+Group Observation	
	8	7	3	4	
Missing Observations	0				
Hypothesis					
H_0 : M = M_0	μ_0	[TestStat]	Rank Sum R+	P(X ≤ R+)	P(X ≥ R+)
H_1 : M ≠ M_0	0.00	+Group Rank Sum(R+)	19.00	0.8125	0.2344

<Figure 10.2.11> Result of Wilcoxon signed rank sum test

- The Wilcoxon signed rank test for the paired samples is to test whether the population median of the differences between two populations, M_d, is zero or not. If we denote the paired samples as $(x_1, y_1), (x_2, y_2), \cdots, (x_n, y_n)$, the Wilcoxon signed rank sum test calculates the difference $d_i = x_i - y_i$ first and assign ranks on $|d_i|$. The sum of ranks of $|d_i|$ which has + sign, R_+, is used as the test statistic. 『eStatU』 provides the distribution of R_+, denoted as $w_+(n)$, up to $n = 22$. $w_+(n)_\alpha$ refers to the right tail 100$\times\alpha$ percentile of this distribution which may not have an accurate percentile value, because it is a discrete distribution. In this case the average of two values near $w_+(n)_\alpha$ is used approximately. Table 10.2.9 summarizes the decision rule of the Wilcoxon signed rank sum test for paired samples by the type of hypothesis.

Table 10.2.9 Wilcoxon signed rank sum test for paired samples

Type of Hypothesis	Decision Rule Test Statistic: R_+ = 'sum of ranks on $\lvert d_i \rvert$ with + sign'
1) $H_0 : M_d = 0$ $H_1 : M_d > 0$	If $R_+ > w_+(n)_\alpha$, then reject H_0, else accept H_0
2) $H_0 : M_d = 0$ $H_1 : M_d < 0$	If $R_+ < w_+(n)_{1-\alpha}$, then reject H_0, else accept H_0
3) $H_0 : M_d = 0$ $H_1 : M_d \neq 0$	If $R_+ < w_+(n)_{1-\alpha/2}$ or $R_+ > w_+(n)_{\alpha/2}$, then reject H_0, else accept H_0

If there is 0 on the differences of paired samples?

If there is 0 on the differences of paired samples, the data is omitted for further analysis. That is, n is decreased.

[Practice 10.2.2]

An oil company has developed a gasoline additive that will improve the fuel mileage of gasoline. We used 8 pairs of cars to compare the fuel mileage to see if it is actually improved. Each pair of cars has the same details as its structure, model, engine size, and other relationship characteristics. When driving the test course using gasoline, one of the pair selected randomly and added additives, the other of the pair was driving the same course using gasoline without additives. The following table shows the km per liter for each of pairs.

pair	Additive (X1)	No Additive (X2)	Difference
1	17.1	16.3	0.8
2	12.7	11.6	1.1
3	11.6	11.2	0.4
4	15.8	14.9	0.9
5	14.0	12.8	1.2
6	17.8	17.1	0.7
7	14.7	13.4	1.3
8	16.3	15.4	0.9

Ex ⇨ eBook ⇨ PR100202_DifferenceOfMileage.csv

Apply a nonparametric test to check whether the additive increase fuel mileage. Use the significance level of 0.05.

- If the sample size of the paired sample is large, use the normal distribution approximation formula shown in Table 10.1.6.

10.3 Nonparametric Test for Location Parameters of Several Populations

- The testing hypothesis for several population means in Chapter 9 was possible if each population could be assumed to be a normal distribution and has the same population variance. However, the assumption that the population follows a normal distribution may not be true for real-world data, or that there may not be enough data to assume a normal distribution. Alternatively, if data are ordinal such as ranks, then the parametric test is not appropriate. In this case, a nonparametric test is used by converting data into ranks without making assumptions about the population distribution. This section introduces the Kruskal-Wallis test corresponding to the completely randomized design of experiments and the Friedman test corresponding to the randomized block design of experiments in Chapter 9.

- Since nonparametric tests are done by using the converted data such as ranks, there may be some loss of information about the data. Therefore, if data are normally distributed, there is no reason to apply a nonparametric test. However, a nonparametric test would be a more appropriate method if data were selected from a population that did not follow a normal distribution.

10.3.1 Completely Randomized Design: Kruskal-Wallis Test

- The Kruskal–Wallis test extends the Wilcoxon rank sum test for two populations. Consider the following example.

Example 10.3.1

The result of a survey of the job satisfaction by sampling employees of three companies are as follows. From this data, can you say that the three companies have different job satisfaction? (unit: points out of 100 scores)

Company A	69	67	65	59
Company B	56	63	55	
Company C	71	72	70	

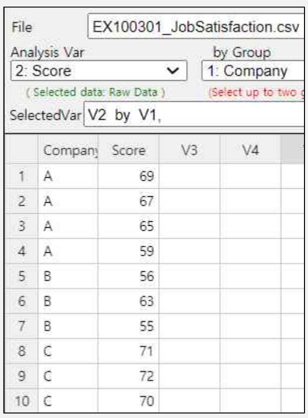
Ex ⇨ eBook ⇨ EX100301_JobSatisfaction.csv

1) Draw a histogram of the data to see whether the comparison of the job satisfaction for the three companies can be made using a parametric test.
2) Using the Kruskal-Wallis test, which is a nonparametric test, find whether the three companies have the same job satisfaction or not with the significance level of 5%
3) Check the above result of the Kruskal-Wallis test using 『eStat』.

Answer

1) The parametric method for testing the hypothesis that three population means are the same is the one-way analysis of variance studied in Chapter 9 and it requires the assumption that the populations are normal distributions. Since the sample sizes are small, n_1=4, n_2=3, n_3=3, in each of the population respectively we need to examine if each sample data satisfy the normality assumption.

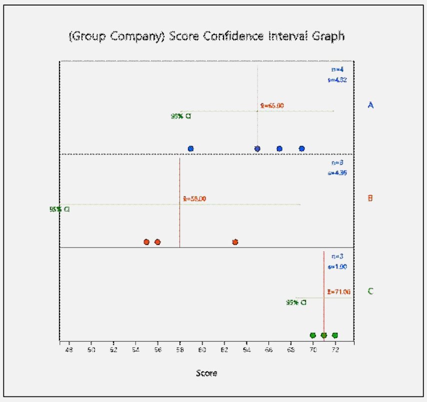
◆ Enter the data as shown in <Figure 10.3.1> in 『eStat』.



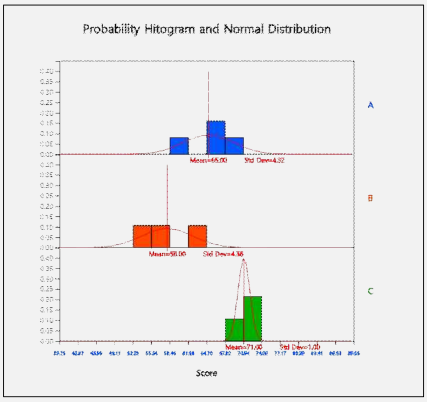
<Figure 10.3.1>
『eStat』 data input

Example 10.3.1 Answer (continued)

- Click the ANOVA icon . Select 'Score' as 'Analysis Var' and 'Company' as 'by Group' variable in the variable selection box. Then a dot graph with the 95% confidence interval of each population mean will appear as in <Figure 10.3.2>. Company C has the highest average of satisfaction scores, followed by Company A and Company B. However, it should be tested if these differences are statistically significant. Clicking the [Histogram] button in the options window below the graph will reveal the histogram and its normal distribution curve for each company, as in<Figure 10.3.3>.



<Figure 10.3.2> Dot graph and the confidence interval by company



<Figure 10.3.3> Histogram by company

- Looking at the histogram, the data are not sufficient to assume that the population follows a normal distribution, because the number of data is so small. In such a case, applying the parametric hypothesis test such as the ANOVA F-test may lead to errors. The hypothesis for this problem is to test whether the location parameters M_1, M_2, M_3 of the three populations are the same or not as follows:

  H_0 : $M_1 = M_2 = M_3$

  H_1 : At least one pair of location parameters is not the same.

- The Kruskal–Wallis test combines all three samples into a single set of data and calculate ranks of this data. If there is a tie, then the average rank will be assigned. Then the sum of the ranks in each sample, R_1, R_2, R_3, is calculated. The test statistic H for the Kruscal–Wallis test is similar to the F-test by converting sample data into ranks as follows:

$$H = \frac{12}{n(n+1)} \sum_{j=1}^{3} \frac{R_j^2}{n_j} - 3(n+1)$$

- To obtain ranks of the combined sample, it is convenient to arrange the data in ascending order separately and then rank the whole data as shown in Table 10.3.1.

Table 10.3.1 A table to calculate the sum of ranks in each sample

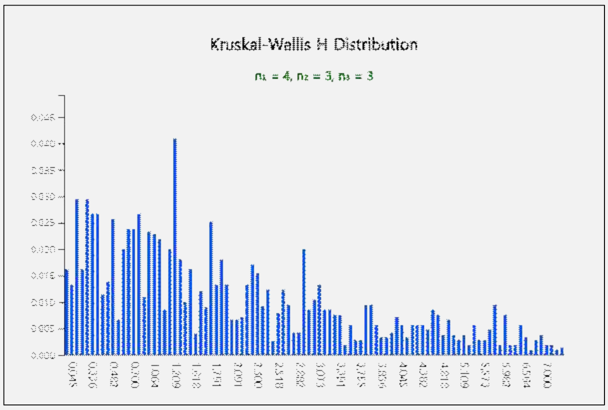
Sample 1 Sorted Data	Sample 2 Sorted Data	Sample 3 Sorted Data	Sample 1 Rank	Sample 2 Rank	Sample 3 Rank
	55			1	
	56			2	
59			3		
	63			4	
65			5		
67			6		
69			7		
		70			8
		71			9
		72			10
		Sum of ranks	R_1 = 21	R_2 = 7	R_3 = 27

Example 10.3.1 Answer (continued)

- The total sum of ranks is $1+2+\cdots+10=10(10+1)/2=55$. The sum of ranks for sample 1 is R_1 = 21, for sample 2 is R_2 = 7, and for sample 3 is R_3 = 27. When the number of data in each sample is taken into account, if R_1, R_2, and R_3 are similar, the null hypothesis that three population location parameters are the same would be accepted. In this example, despite of the small sample size for sample 3, R_3 is larger thant R_1 or R_2. Also R_1 is larger than R_2. Based on these differences, can you conclude that the three population location parameters are statistically different?
- In the above example, the H statistic is as follows:

$$H = \frac{12}{10(10+1)} \left(\frac{21^2}{4}+\frac{7^2}{3}+\frac{27^2}{3}\right) - 3(10+1) = 7.318$$

If the null hypothesis is true, the distribution of the test statistic should be known to investigate how large a value of H is statistically significant. If $n=10$, the number of cases for ranking {1,2,3, $\cdots$, 10} is $10! = 3{,}628{,}800$. It is not easy to examine all of these possible rankings to create a distribution table of H. 『eStat U』 shows the distribution of the Kruskal–Wallis H for n_1=4, n_2=3, and n_3=3 as shown in <Figure 10.3.4>, and a part of the distribution table as in Table 10.3.2. As shown in the figure, the distribution of H is an asymmetrical distribution.



<Figure 10.3.4> Kruskal Wallis H distribution when n_1=4, n_2=3, n_3=3

Table 10.3.2 Kruskal Wallis H distribution when n_1=4, n_2=3, n_3=3

Kruskal Wallis H distribution	k = 3		
	n_1 = 4	n_2 = 3	n_3 = 3
x	P(X = x)	P(X $\le$ x)	P(X $\ge$ x)
0.018	0.0162	0.0162	1.0000
0.045	0.0133	0.0295	0.9838
$\cdots$	$\cdots$	$\cdots$	$\cdots$
5.727	0.0048	0.9543	0.0505
5.791	0.0095	0.9638	0.0457
5.936	0.0019	0.9657	0.0362
5.982	0.0076	0.9733	0.0343
6.018	0.0019	0.9752	0.0267
6.155	0.0019	0.9771	0.0248
6.300	0.0057	0.9829	0.0229
6.564	0.0033	0.9862	0.0171
6.664	0.0010	0.9871	0.0138
6.709	0.0029	0.9900	0.0129
6.745	0.0038	0.9938	0.0100
7.000	0.0019	0.9957	0.0062
7.318	0.0019	0.9976	0.0043
7.436	0.0010	0.9986	0.0024
8.018	0.0014	1.0000	0.0014

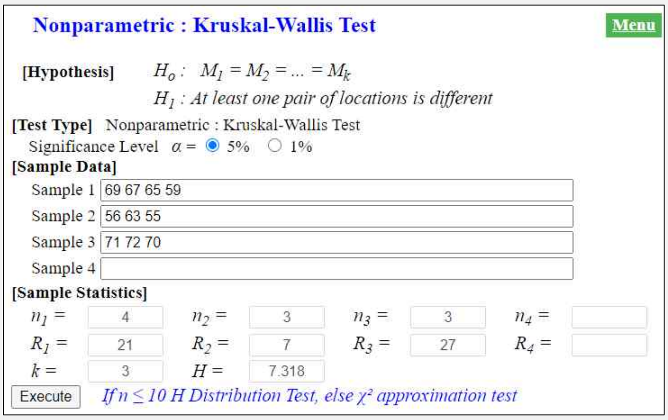
Example 10.3.1 Answer (continued)

- H test is a right tail test and the 5 percentile from the right corresponding to the significance level is approximately P(X $\geq$ 5.727) = 0.0505. Note that there is no exact 5 percentile in case of a discrete distribution. Hence, the decision rule to test the null hypothesis is as follows:

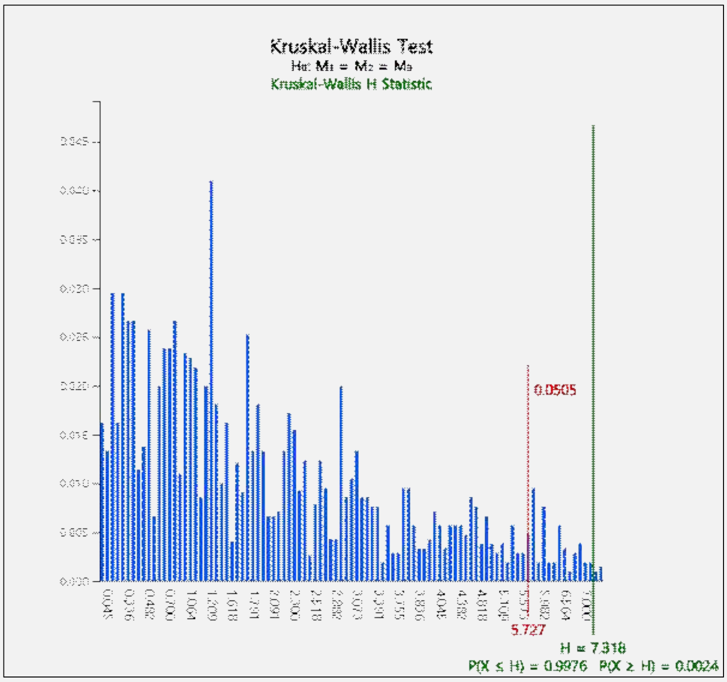
  'If $H > 5.727$, then reject H_0'

  Since H = 7.318 in this example, we reject H_0.

3) In 『eStatU』, enter data as <Figure 10.3.5> and click the [Execute] button. Then the sample statistics are calculated and the test result is shown as in <Figure 10.3.6>. The critical line for values containing 5 percentile of the significance level is shown here. For a discrete distribution, the choice of the final rejection region shall be determined by the analyst.



<Figure 10.3.5> 『eStatU』 Kruskal-Wallis test



<Figure 10.3.6> Kruskal-Wallis test

Example 10.3.1 Answer (continued)

- 『eStat』 may also be used to conduct the Kruskal–Wallis H test. Enter data as <Figure 10.3.1> and click the ANOVA icon. Select 'Score' as 'Analysis Var' and 'Company' as 'by Group' variable in the variable selection box. Then a dot graph with the 95% confidence interval of the population mean in each company will appear as <Figure 10.3.2>. If you press the [Kruskal–Wallis test] button in the options window below the graph, the same test graph and test result table will appear as in <Figure 10.3.7>.

Kruskal-Wallis Test	Analysis Var	Score		
Statistics	Observation	Mean	Std Dev	Rank Sum
1 (A)	4	65.000	4.320	21.00
2 (B)	3	58.000	4.359	7.00
3 (C)	3	71.000	1.000	27.00
Total	10	64.700	6.237	55.00
Missing Observations	0			
Hypothesis				
$H_0 : M_1 = M_2 = M_3$	[TestStat]	H	P(X ≤ H)	P(X ≥ H)
At least one pair of locations is different	H	7.318	0.9976	0.0024

<Figure 10.3.7> Result of the Kruskal-Wallis test

- Let us generalize the Kruskal–Wallis H test described so far with an example. Denote random samples collected independently from the k populations (at each level of one factor) when their sample sizes are n_1, n_2, ..., n_k as follows: $(n = n_1 + n_2 + \cdots + n_k)$.

Table 10.3.3 Notation for random samples from each level

Level 1	Level 2	$\cdots$	Level k	
X_{11} X_{12} $\cdots$ X_{1n_1}	X_{21} X_{22} $\cdots$ X_{2n_2}	$\cdots$	X_{k1} X_{k2} $\cdots$ X_{kn_k}	
Level 1 Mean $\overline{X}_{1}\cdot$	Level 2 Mean $\overline{X}_{2}\cdot$	$\cdots$	Level k Mean $\overline{X}_{k}\cdot$	Total Mean $\overline{X}\cdot\cdot$

- The statistical model of the Kruskal-Wallis test is as follows:

$$X_{ij} = \mu + \tau_i + \epsilon_{ij}, \ i = 1,2,\cdots,k; \ j = 1,2,\cdots,n_i \quad \text{where} \quad \sum_{i=1}^{k} \tau_i = 0.$$

Here τ_i represents the effect of the level i and ϵ_{ij}'s are independent and follow the same continuous distribution.

- The hypothesis of the Kruskal-Wallis test is as follows:

$H_0 : \tau_1 = \tau_2 = \cdots = \tau_k$
H_1 : At least one pair of τ_j is not equal.

- For the Kruskal-Wallis test, ranking data for the combined sample must be created. Table 10.3.4 is a notation of ranking data for each level.

Table 10.3.4 Notation of ranking data in each level

Level 1	Level 2	$\cdots$	Level k	
R_{11} R_{12} $\cdots$ R_{1n_1}	R_{21} R_{22} $\cdots$ R_{2n_2}	$\cdots$	R_{k1} R_{k2} $\cdots$ R_{kn_k}	
Sum of ranks in level 1 $R_{1\cdot}$	Sum of ranks in level 2 $R_{2\cdot}$	$\cdots$	Sum of ranks in level k $R_{k\cdot}$	
Mean of ranks in level 1 $\overline{R}_{1\cdot}$	Mean of ranks in level 2 $\overline{R}_{2\cdot}$	$\cdots$	Mean of ranks in level k $\overline{R}_{k\cdot}$	Total mean of ranks $\overline{R}_{\cdot\cdot} = (n+1)/2$

- The sum of squares for the one-way analysis of variance studied in Chapter 9 by using the ranking data in Table 10.3.4 are as follows:

$$SST = \sum_{i=1}^{k}\sum_{j=1}^{n_i}(R_{ij} - \overline{R}_{\cdot\cdot})^2 = \sum_{k=1}^{n}(k - \overline{R}_{\cdot\cdot})^2 = n(n+1)(n-1)$$
$$SSTr = \sum_{i=1}^{k}\sum_{j=1}^{n_i}(\overline{R}_{i\cdot} - \overline{R}_{\cdot\cdot})^2$$
$$SSE = SST - SSTr$$

- Also, the statistic for the F-test is as follows:

$$F = \frac{MSTr}{MSE} = \frac{\frac{SSTr}{k-1}}{\frac{SSE}{n-k}} = \frac{\frac{SSTr}{k-1}}{\frac{SST-SSTr}{n-k}} = \frac{\frac{n-k}{k-1}}{\frac{SST}{SSTr}-1}$$

Since SST is a constant, the statistic for the F-test is proportional to $SSTr$.

- The statistic for the Kruskal-Wallis test H is proportional to $SSTr$ as follows:

$$H = \frac{12}{n(n+1)} \sum_{i=1}^{k} n_i(\overline{R}_{i\cdot} - \overline{R}_{\cdot\cdot})^2$$
$$= \frac{12}{n(n+1)} \sum_{i=1}^{k} R_{i\cdot}^2 - 3(n+1)$$

- The multiplication constant $\frac{12}{n(n+1)}$ in the definition of H statistics is intended to ensure that the statistic follows approximately the chi-square distribution with $k-1$ degrees of freedom.
- The distribution of the Kruskal-Wallis test statistic H, denoted as $h(n_1,n_2,\cdots,n_k)$, can be obtained by considering all possible cases of ranks {1, 2, $\cdots$, n} which is $n!$. 『eStatU』 provides the table of $h(n_1,n_2,\cdots,n_k)$ up to $n=10$. $h(n_1,n_2,\cdots,n_k)_\alpha$ denotes the right tail 100$\times\alpha$ percentile, but it might not have the exact value of this percentile, because $h(n_1,n_2,\cdots,n_k)$ is a discrete distribution. In this case, the middle of two adjacent values of 100$\times\alpha$ percentile is often used. The decision rule of the Kruskal-Wallis test is as Table 10.3.5.

Table 10.3.5 Kruskal-Wallis test

Hypothesis	Decision Rule Test Statistic: H
$H_0 : \tau_1 = \tau_2 = \cdots = \tau_k$ H_1 : At least one pair of τ_j is not equal.	If $H > h(n_1, n_2, \cdots, n_k)_\alpha$, then reject H_0, else accept H_0.

☞ **If there are tied values in the combined sample, assign the average of ranks.**

- The distribution of the Kruskal-Wallis H statistic is independent of a population distribution. In other words, the Kruskal-Wallis test is a distribution-free test.
- If the null hypothesis is true and the sample size is large enough, the test statistic H is approximated by the chi-square distribution with $k-1$ degrees of freedom. Table 10.3.6 summarizes the decision rule for the Kruskal-Wallis test in case of large samples.

Table 10.3.6 Kruskal-Wallis test in case of large samples.

Hypothesis	Decision Rule Test Statistic: H
$H_0 : \tau_1 = \tau_2 = \cdots = \tau_k$ H_1 : At least one pair of τ_j is not equal.	If $H > \chi^2_{k-1\,;\,\alpha}$, then reject H_0, else accept H_0

- If there is a tie in the combined sample, the average rank is assigned to each data. In this case, the statistic H shall be modified as follows:

$$H' = \frac{H}{1 - \sum_{j=1}^{g} \frac{T_j}{n^3 - n}}$$

Here g = (number of tied groups)

$$T_j = \sum_{j=1}^{g} t_j (t_j - 1)(t_j + 1)$$

t_j = (the size of the j^{th} tie group, i.e., the number of observations in the tie group)
if there is no tie, the size of the j^{th} tie group is 1 and t_j=1.

[Practice 10.3.1]

A bread maker wants to compare the three new mixing methods of ingredients. 15 breads were made by each mixing method (A, B, C) of 5 pieces, and a group of judges who did not know the difference in material mixing ratio gave the following points. Test the null hypothesis that there is no difference in taste according to the mixing methods at the significance level of 0.05.

Mixing ratio:	Method A:	72	88	70	87	71
	Method B:	85	89	86	82	90
	Method C:	94	94	88	87	89

Ex ⇨ eBook ⇨ PR100301_ScoreByMixingMethod.csv

10.3.2 Randomized Block Design: Friedman Test

- In Section 9.2, we studied the randomized block design to measure the fuel mileage of three types of cars which reduce the impact of the block factor, i.e., driver. If each population follows a normal distribution, sample data are analyzed using the F-test based on the two-way analysis of variance without the interaction. However, the assumption that a population follows a normal distribution may not be appropriate for real-world data, or that there may not be enough data to assume a normal distribution. Alternatively, if the data collected might not be continuous and are ordinal such as ranks, then the parametric test is not appropriate. In such cases, nonparametric tests are used to test parameters by converting data to ranks without assuming the distribution of the population. This section introduces the Friedman test corresponding to the randomized block design experiments in Section 9.2.2.
- Let us take a look at the Friedman test using [Example 9.2.1] which was the car fuel mileage measurement problem.

Example 10.3.2

The fuel mileage of the three types of cars (A, B and C) is measured using the randomized block design as Table 9.2.4 and it is rearranged in Table 10.3.7.

Table 10.3.7 Fuel mileage of the three types of cars

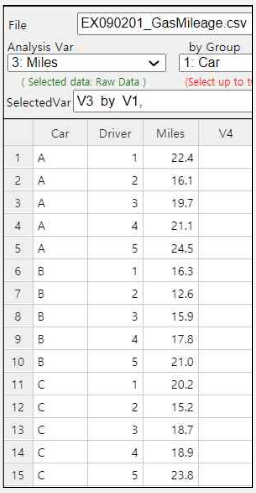
		Car A	Car B	Car C
Driver (Block)	1	22.4	16.3	20.2
	2	16.1	12.6	15.2
	3	19.7	15.9	18.7
	4	21.1	17.8	18.9
	5	24.5	21.0	23.8

Ex ⇨ eBook ⇨ EX090201_GasMileage.csv

1) Draw a histogram of the data to see if the fuel mileage of the three cars can be tested by a parametric method.
2) Using the Friedman test which is a nonparametric method of the randomized block design, test whether the fuel mileage of the three types of cars are different with the significance level of 5%.
3) Check the result of the above Friedman test using 『eStatU』.

Answer

1) Enter data in 『eStat』 as shown in <Figure 10.3.8>.

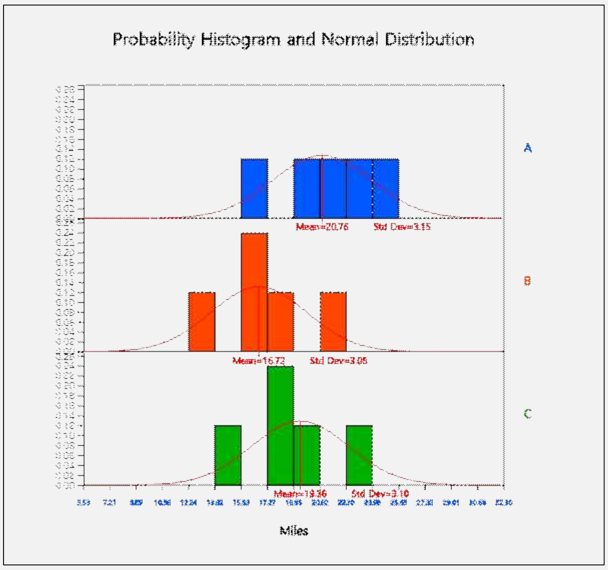


	Car	Driver	Miles	V4
1	A	1	22.4	
2	A	2	16.1	
3	A	3	19.7	
4	A	4	21.1	
5	A	5	24.5	
6	B	1	16.3	
7	B	2	12.6	
8	B	3	15.9	
9	B	4	17.8	
10	B	5	21.0	
11	C	1	20.2	
12	C	2	15.2	
13	C	3	18.7	
14	C	4	18.9	
15	C	5	23.8	

<Figure 10.3.8>
『eStat』 Data input

Example 10.3.2 Answer (continued)	

- Click icon  of the analysis of variance. Select ‘Miles’ as 'Analysis Var' and ‘Car’ as ‘by Group’. Then the dot graph by car type and the 95% confidence interval for the population mean will appear. Again, clicking the [Histogram] button in the options window below the graph will show the histogram and normal distribution curve for each car type as shown in <Figure 10.3.9>.



<Figure 10.3.9> Histogram of fuel mileage by car

- Looking at the histogram, it is not sufficient to assume that each population follows a normal distribution, because of the small number of data. In such case, applying the parametric F-test may lead to errors.
- The hypothesis for this problem is to test whether or not the location parameters M_1, M_2, M_3 of the three populations are the same.

H_0 : $M_1 = M_2 = M_3$
H_1 : At least one pair of location parameters is not equal.

- The Friedman test calculates the sum of ranks, R_1, R_2, R_3, for each of the three types of cars after the ranking is calculated for the fuel mileage measured for each driver (block) (Table 10.3.8). If there is a tie, then the average of ranks is assigned.

Table 10.3.8 Ranking in each of the block

		Car A	Car B	Car C
Driver (Block)	1	3	1	2
	2	3	1	2
	3	3	1	2
	4	3	1	2
	5	3	1	2
Sum of ranks		R_1=15	R_2=5	R_3=10

The sum of ranks for Car A is R_1 = 15, for Car B is R_2 = 5, for Car C is R_3 = 10. The sum of ranks looks different. Are the differences statistically significant?

- The Friedman test statistic S can be considered as the F statistic in the two-way analysis of variance to these ranking data as follows:

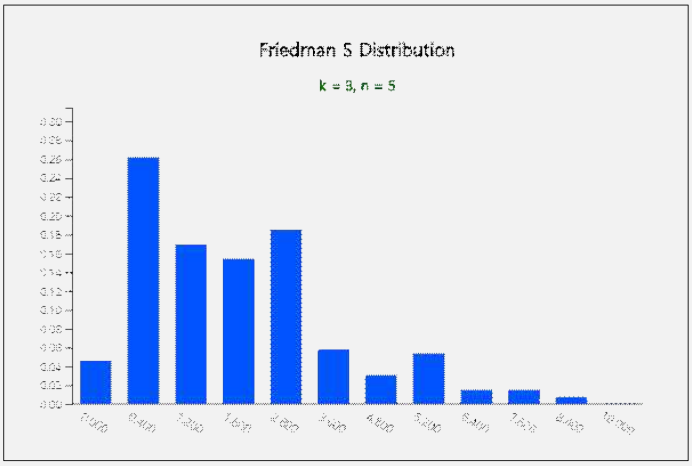
$$S = \frac{12}{nk(k+1)} \sum_{j=1}^{k} R_j^2 - 3n(k+1),$$ k is the number of population.

Example 10.3.2 Answer (continued)

In this example, $k=3$ and the S statistic is as follows:

$$S = \frac{12}{5\times 3(3+1)}\left(15^2+5^2+10^2\right) - 3\times 5(3+1) \;=\; 10$$

The distribution of the test statistic S, when the null hypothesis is true, should be known to investigate how large a value of S is statistically significant. Since the number of cases of ranking when $n=5,\ k=3$ is $(3!)^5 = 7776$, it is not easy to examine all of these possible rankings to obtain a distribution. 『eStatU』 provides the distribution of the test statistic S in the case of $n=5,\ k=3$ as in <Figure 10.3.10> and its distribution table as Table 10.3.9. As shown in the graph, the distribution of S is an asymmetrical distribution.



<Figure 10.3.10> Friedman S distribution when $n=5,\ k=3$

Table 10.3.9 Friedman S distribution when $n=5,\ k=3$

Friedman S distribution	k = 3	n = 5	
x	P(X = x)	P(X $\le$ x)	P(X $\ge$ x)
0.000	0.0463	0.0463	1.0000
0.400	0.2623	0.3086	0.9537
1.200	0.1698	0.4784	0.6914
1.600	0.1543	0.6327	0.5216
2.800	0.1852	0.8179	0.3673
3.600	0.0579	0.8758	0.1821
4.800	0.0309	0.9066	0.1242
5.200	0.0540	0.9606	0.0934
6.400	0.0154	0.9761	0.0394
7.600	0.0154	0.9915	0.0239
8.400	0.0077	0.9992	0.0085
10.000	0.0008	1.0000	0.0008

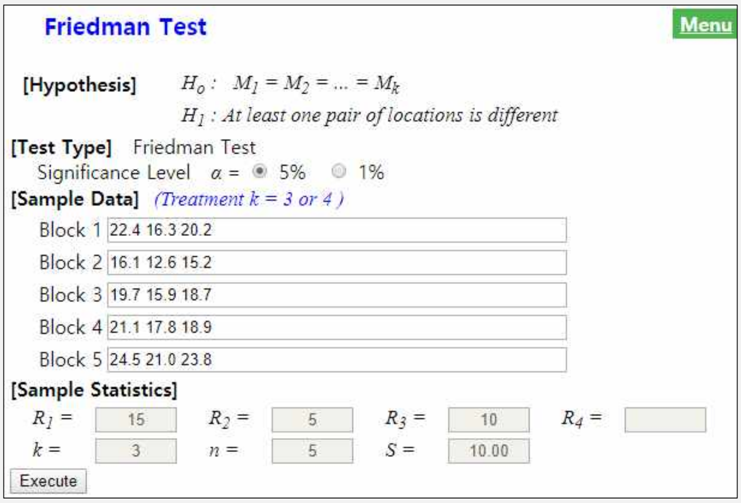
- The Friedman test is a right sided test. If we look for the five percentile from the right tail corresponding to significance level, the nearest value is P(X $\ge$ 6.4) = 0.0394. Since it is a discrete distribution, there is no exact value of five percentile. Hence, the rejection region with the significance level of 5% can be written as follows:

  ‘If $S \ge 6.4$, then reject H_0’

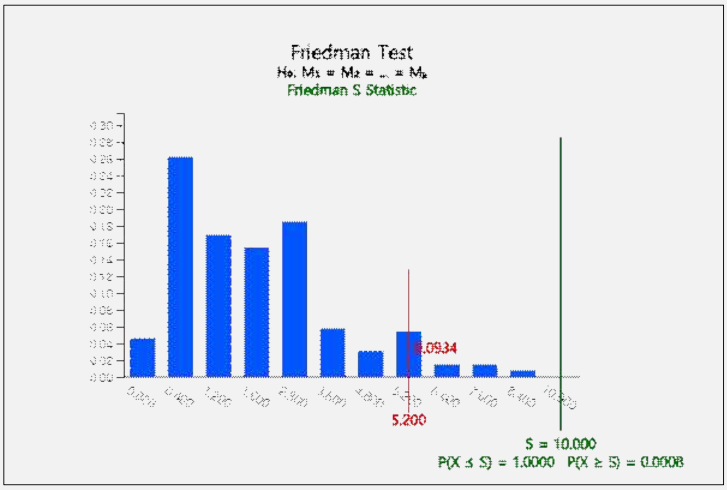
Since S = 10 in this example, H_0 is rejected.

Example 10.3.2 Answer (continued)

3) Enter data in『eStatU』as in <Figure 10.3.11> and click the [Execute] button. The sample statistics and test graph will be shown as in <Figure 10.3.12>. The critical line which contains 5% of the significance level is shown here. For a discrete distribution, the choice of the final rejection region should be determined by the analyst.



<Figure 10.3.11> Data input for Friedman test at『eStatU』



<Figure 10.3.12> Result of Friedman test using『eStatU』

- Let's generalize the Friedman test described so far using the above example. Assume that there are k number of levels and denote the rank of n number of data as follows:

Table 10.3.10 Notation of n random samples for k number of levels with randomized block design

Block \ Treatment	Level 1	Level 2	$\cdots$	Level k	
1 2 $\cdots$ n	X_{11} X_{12} $\cdots$ X_{1n}	X_{21} X_{22} $\cdots$ X_{2n}	$\cdots$	X_{k1} X_{k2} $\cdots$ X_{kn}	
Mean	$\overline{X}_{1}\cdot$	$\overline{X}_{2}\cdot$	$\cdots$	$\overline{X}_{k}\cdot$	Total Mean $\overline{X}\cdot\cdot$

- A statistical model of the Friedman test is as follows:

  $$X_{ij} = \mu + \tau_i + \beta_j + \epsilon_{ij}, \qquad i = 1,2,\cdots,k;\ j = 1,2,\cdots,n$$

  Here τ_i is the effect of level i which satisfies $\sum_{i=1}^{k} \tau_i = 0$ and β_j is the effect of block j which satisfies $\sum_{j=1}^{n} \beta_j = 0$. ϵ_{ij}'s are independent and follows the same continuous distribution.
- The hypothesis of the Friedman test is as follows:

  H_0 : $\tau_1 = \tau_2 = \cdots = \tau_k$
  H_1 : At least one pair of τ_i is different

- For the Friedman test, ranking data for each block must be created. Table 10.3.11 is the notation of ranking data for each level.

Table 10.3.11 Notation of rank data in each level

Block \ Treatment	Level 1	Level 2	$\cdots$	Level k	
1 2 $\cdots$ n	R_{11} R_{12} $\cdots$ R_{1n}	R_{21} R_{22} $\cdots$ R_{2n}	$\cdots$	R_{k1} R_{k2} $\cdots$ R_{kn}	
Sum of ranks	$R_{1}\cdot$	$R_{2}\cdot$	$\cdots$	$R_{k}\cdot$	
Average of ranks	$\overline{R}_{1}\cdot$	$\overline{R}_{2}\cdot$	$\cdots$	$\overline{R}_{k}\cdot$	Average of ranks $\overline{R}\cdot\cdot = (k+1)/2$

- If we apply the analysis of variance for the rank data of Table 10.3.11 instead of the observation data in Section 9.2, the total sum of squares, SST, and the block sum of squares SSB are constants. The treatment sum of squares $SSTr$ is as follows:

  $$SSTr = \sum_{i=1}^{k} n(\overline{R}_{i\cdot} - \overline{R}\cdot\cdot)^2$$
  $$SST = SSTr + SSE$$

  Therefore, the F test statistic can be written as follows:

$$F = \frac{MSTr}{MSE} = c\frac{SSTr}{SST - SSTr - SSE},$$ Here c is a constant.

That is, since SST is a constant, F test statistic is proportional to $SSTr$.

- The Friedman test statistic S is proportional to $SSTr$ as follows:

$$S = \frac{12}{k(k+1)} SSTr = \frac{12n}{k(k+1)} \sum_{i=1}^{k} (\overline{R}_{i\cdot} - \overline{R}_{\cdot\cdot})^2$$
$$= \frac{12}{nk(k+1)} \sum_{i=1}^{k} R_{i\cdot}^2 - 3n(k+1)$$

The reason S statistic has the constant multiplication of $\frac{12}{k(k+1)}$ is to make S which follows a chi-square distribution with $k-1$ degrees of freedom.

- The distribution of the Friedman test statistic S is denoted as $s(k,n)$. 『eStatU』 provides the distribution of $s(k,n)$ up to $n \le 8$ if $k=3$ and up to $n \le 6$ if $k=4$. $s(k,n)_\alpha$ denotes the right tail $100\times\alpha$ percentile, but there might not be the exact percentile, because it is a discrete distribution. In this case, the middle value of two nearest $s(k,n)_\alpha$ is often used approximately. Table 10.3.12 is the summary of decision rule of the Friedman test.

Table 10.3.12 Friedman Test

Hypothesis	Decision Rule Test Statistic: S
$H_0: \tau_1 = \tau_2 = \cdots = \tau_k$ H_1 : At least one pair of τ_j is different	If $S > s(k,n)_\alpha$, then reject H_0, else accept H_0

☞ **If there are tied values on each block, use the average rank.**

- The distribution of the Friedman statistic S is independent of the population distribution. In other words, the Friedman test is a distribution-free test.
- If the null hypothesis is true and if the sample is large enough, the test statistic S is approximated by the chi-square distribution with $k-1$ degrees of freedom. Table 10.3.13 summarizes the decision rule for the Friedman test in case of large sample.

Table 10.3.13 Friedman Test – large sample case

Hypothesis	Decision Rule Test Statistic: S
$H_0: \tau_1 = \tau_2 = \cdots = \tau_k$ H_1 : At least one pair of τ_j is different	If $S > \chi^2_{k-1\,;\,\alpha}$, then reject H_0, else accept H_0

- If there is a tie in the block, the average rank is assigned to each data. In this case, the statistic S shall be modified as follows:

$$S' = \frac{S}{1-\sum_{j=1}^{g}\frac{T_j}{np(p^2-1)}}$$

Here g = (number of tied groups)

$$T_j = \sum_{j=1}^{g} t_j(t_j-1)(t_j+1)$$

t_j = (the size of the j^{th} tie group, i.e., the number of observations in the tie group).
If there is no tie, the size of the j^{th} tie group is 1 and t_j=1

[Practice 10.3.2]

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). Apply the Friedman test whether the mean yields of the four wheats are the same or not with the 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 ⇨ PR100302_WheatAreaYield.csv

Exercise

10.1 A psychologist has selected 12 handicap workers randomly from production workers employed at various factories in a large industrial complex and their work competency scores are examined as follows. The psychologist wants to test whether the population average score is 45 or not. Assume the population distribution is symmetrical about the mean.

32, 52, 21, 39, 23, 55, 36, 27, 37, 41, 34, 51

1) Check whether a parametric test is possible.
2) Apply the sign test with the significance level of 5%.
3) Apply the Wilcoxon signed rank test with the significance level of 5%.

10.2 A tire production company wants to test whether a new manufacturing process can produce a more durable tire than the existing process. The tire by a new process was tested to obtain the following data: (unit: 1000km)

Existing Process	New Process
62	73
76	53
61	61
90	65
74	60
74	53
75	70
63	63

1) Check whether a parametric test is possible.
2) Apply the Wilcoxon rank sum test whether the new process and the existing process have the same durability or not with the significance level of 5%.

10.3 A company wants to compare two methods of obtaining information about a new product. Among company employees, 19 were randomly selected and divided into two groups. The first group learned about the new product by the method A, and the second group learned by the method B. At the end of the experiment, the employees took a test to measure their knowledge of the new product and their test scores are as follows. Can we conclude from these data that the median values of the two groups are different? Test with the significance level of 0.05.

Method A	50	59	60	71	80	78	72	77	73	75
Method B	52	54	58	78	65	61	60	72	60	

10.4 10 men and 10 women working in the same profession were selected independently and their monthly salaries were surveyed. Can you say that a man in this profession earns more than a woman. Test with the significance level of 0.05. (Unit: 10USD)

Man	Woman
381	284
294	279
296	288
389	383
281	489
194	287
193	496
286	393
384	277
494	371

10.5 To find out the fuel mileage improvement effect of a new gasoline additive, 10 cars of the same state were selected. The gas mileage was tested without gasoline additives and with additives running the same road at the same speed and obtain the following data. Test whether the new gasoline additive is effective in improving the fuel mileage with the significance level of 0.05.

gas mileage (unit: km/liter)

With additives	Without additives
11.7	10.3
13.8	12.9
11.2	12.5
7.7	9.5
8.2	11.2
16.3	14.6
14.2	15.9
19.4	18.5
13.9	12.0
15.5	15.1

10.6 In order to determine the efficacy of the new pain reliever, seven persons were tested with aspirin and new pain reliever. The experimental time of the two pain relievers were sufficiently spaced, and the order of the medication experiment was randomly determined. The time (in minutes) until feeling pain relief was measured as follows. Do the data indicate that the new pain reliever has faster pain relief than aspirin? Test with the significance level of 0.05.

Person ID	1	2	3	4	5	6	7
Aspirin	15	20	12	20	17	14	17
New pain reliever	7	14	13	11	10	16	11

10.7 A person was asked to taste 15 coffee samples to rank from 1 (hate first) to 15 (best). The 15 samples are taken from each of the three types of coffee (A, B, C) and are tasted in random order. The following table shows the ranking of preference by the coffee type. Test the null hypothesis that there is no difference in three types of coffee preferences at the significance level of 0.05.

Coffee Type	Ranking				
A	9	10	11	12	13
B	14	1	5	7	8
C	2	3	4	15	6

10.8 A bread maker wants to compare the four new mix of ingredients. 5 breads were made by each mixing ratio of ingredients, a total of 20 breads, and a group of judges who did not know the difference in mixing ratio of ingredients were given the following points. Test the null hypothesis that there is no difference in taste according to the mixing ratio of ingredients at the significance level of 0.05.

Mixing Ratio	Scores				
Method A	72	88	70	87	71
Method B	85	89	86	82	90
Method C	94	94	88	87	89
Method D	91	93	92	95	96

Multiple Choice Exercise

10.1 What is NOT the reason to have a nonparametric test?

① Population is not normally distributed.
② Ordinal data.
③ Data follows a normal distribution.
④ There is an extreme point in sample.

10.2 Which of the following nonparametric tests is for testing the location parameter of single population?

① Wilcoxon signed rank sum test ② Wilcoxon rank sum test
③ Kruskal-Wallis test ④ Friedman test

10.3 Which of the following nonparametric tests is for testing the location parameters of two populations?

① Wilcoxon signed rank sum test ② Wilcoxon rank sum test
③ Kruskal-Wallis test ④ Friedman test

10.4 Which of the following nonparametric tests is for tesing the location parameters of multiple populations?

① Wilcoxon signed rank sum test ② Wilcoxon rank sum test
③ Kruskal-Wallis test ④ Friedman test

10.5 Which of the following nonparametric tests is appropriate for testing of the randomized block design method?

① Wilcoxon signed rank sum test ② Wilcoxon rank sum test
③ Kruskal-Wallis test ④ Friedman test

10.6 What is the sign test?

① Test for the location parameter of single population
② Test for two location parameters of two populations
③ Test for several location parameters of multiple populations
④ Test for the randomized block design

10.7 What is the transformation of data that is often used for nonparametric tests?

① log transformation ② exponential transformation
③ (0-1) transformation ④ ranking transformation

10.8 What is the test statistic used for the sign test?

① rank ② (number of + signs) - (number of - signs)
③ degrees of freedom ④ (number of + signs)

10.9 What is the test statistic used for testing two location parameters of two populations using a nonparametric test?

① (number of + signs)
② sum of ranks in population 2
③ (number of - signs)
④ (sum of ranks in population 1) + (sum of ranks in population 2)

10.10 What is the theoretical basis for the H statistic used for the Kruskal-Wallis test?

① Within sum of squares of rank data
② Error sum of squares of rank data
③ Total sum of squares of rank data
④ Treatment sum of squares of rank data

(Answers)
10.1 ③, 10.2 ①, 10.3 ②, 10.4 ③, 10.5 ④, 10.6 ①, 10.7 ④, 10.8 ④, 10.9 ②, 10.10 ④,