Introduction to Statistics and Data Science using *eStat*

Chapter 10 Nonparametric Testing Hypothesis

10.3 Nonparametric Test for Comparing Locations of Several Populations
10.3.1 Kruskal-Wallis Test

Jung Jin Lee
Professor of Soongsil University, Korea
Visiting Professor of ADA University, Azerbaijan

10.1 Nonparametric Test for Location of Single Population
- 10.1.1 Sign Test
- 10.1.2 Wilcoxon Signed Rank Sum Test

10.2 Nonparametric Test for Comparing Locations of Two Populations
- 10.2.1 Independent Samples: Wilcoxon Rank Sum Test
- 10.2.2 Paired Samples: Wilcoxon Signed Rank Sum Test

10.3 Nonparametric Test for Comparing Locations of Several Populations
- 10.3.1 Completely Randomized Design: Kruskal-Wallis Test
- 10.3.2 Randomized block design: Friedman Test

10.3 Nonparametric Test for Location of Several Populations

10.3.1 Completely Randomized Design : Kruskal-Wallis Test

[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?

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

1) Draw a histogram of the data to see whether the comparison of job satisfaction for three companies can be made using a parametric test.
2) Using the Kruskal-Wallis test, test 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』.

10.3 Nonparametric Test for Location of Several Populations

<Answer of Example 10.3.1>

File EX100301_JobSatisfaction.csv

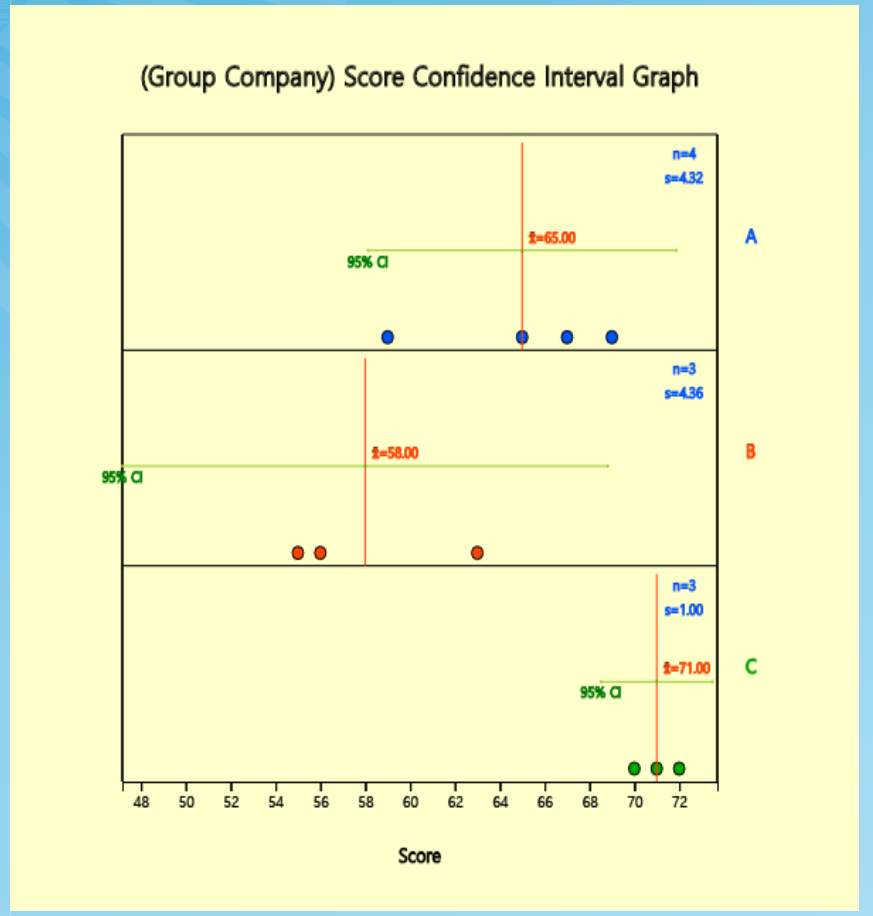
Analysis Var: 2: Score

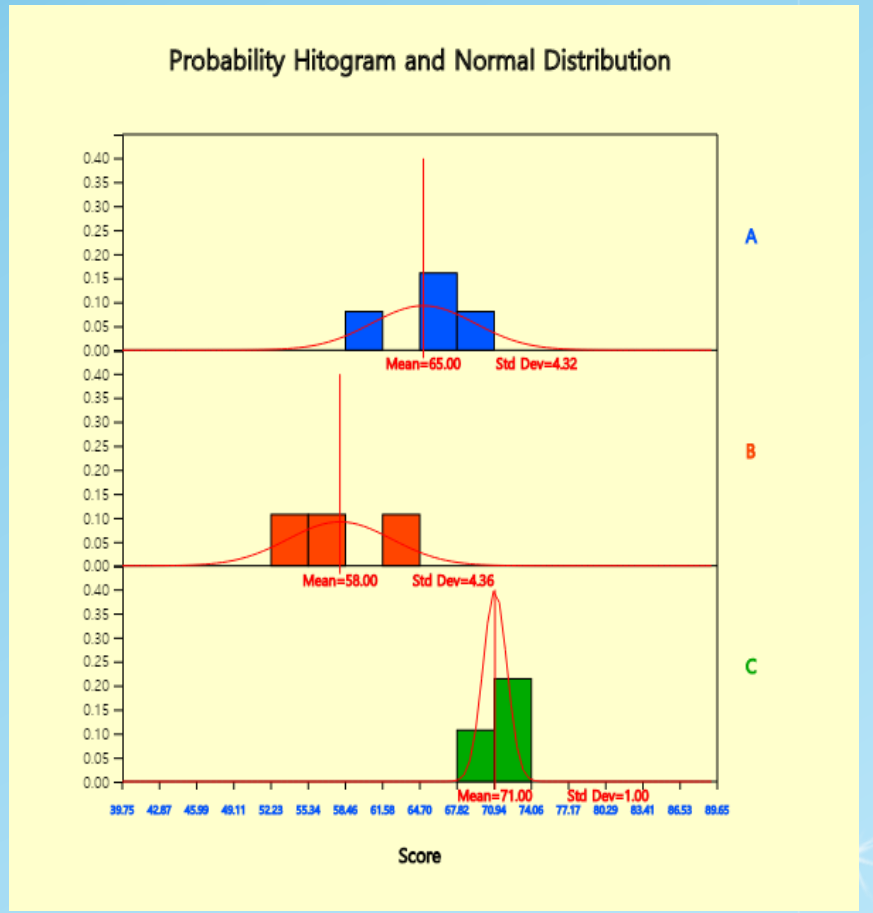
by Group: 1: Company

(Selected data: Raw Data) (Select up to two g

SelectedVar: V2 by V1,

	Company	Score	V3	V4
1	A	69		
2	A	67		
3	A	65		
4	A	59		
5	B	56		
6	B	63		
7	B	55		
8	C	71		
9	C	72		
10	C	70		





10.3 Nonparametric Test for Location of Several Populations

<Answer of Example 10.3.1>

- Hypothesis $H_o : M_1 = M_2 = M_3$

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

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	
			5		
65			6		
67			7		
69					
		70			8
		71			9
		72			10
		Sum of ranks	$R_1 = 21$	$R_2 = 7$	$R_3 = 27$

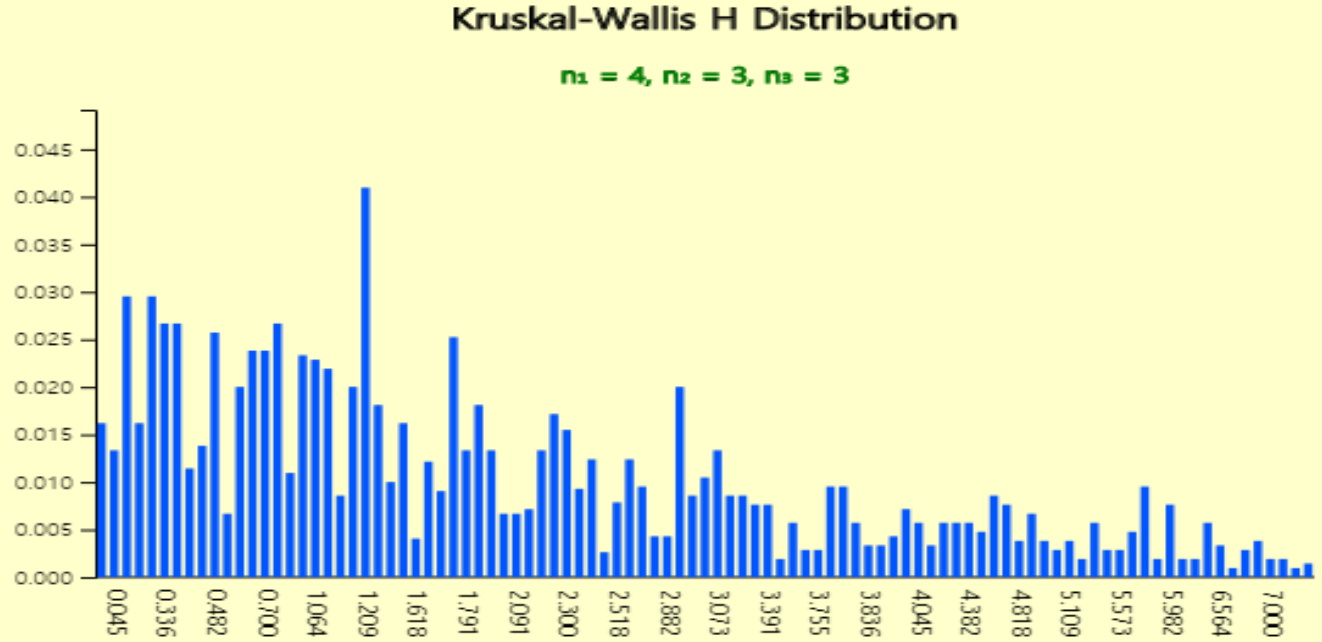
10.3 Nonparametric Test for Location of Several Populations

- **Kruskal-Wallis Test Statistic**

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

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

- **If H > 5.727, then reject H_0 , hence reject H_0**



H distribution	K = 3		
	n_1 = 4	n_2 = 3	n_3 = 3
x	P(X = x)	P(X x)	P(X x)
0.018	0.0162	0.0162	1.0000
0.045	0.0133	0.0295	0.9838
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

10.3 Nonparametric Test for Location of Several Populations

<Answer of Example 10.3.1>

Testing Hypothesis ANOVA

Menu

[Hypothesis] $H_o: \mu_1 = \mu_2 = \ldots = \mu_k$

H_1 : *At least one pair of means is different*

[Test Type] *F test* (ANOVA)

Significance Level α = (•) 5% ( ) 1%

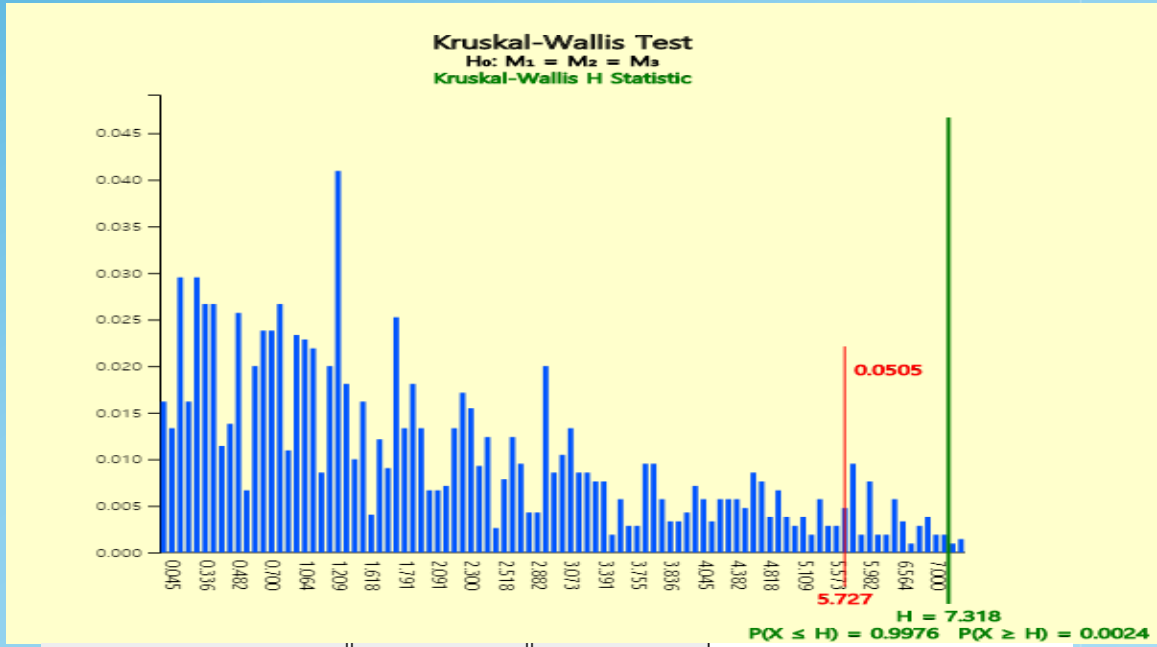
[Sample Data] *Input either sample data using BSV or sample statistics at the next boxes*

Sample 1	69 67 65 59
Sample 2	56 63 55
Sample 3	71 72 70
Sample 4	

[Sample Statistics]

n_1 =	4	n_2 =	3	n_3 =	3	n_4 =	
m_1 =	65.00	m_2 =	58.00	m_3 =	71.00	m_4 =	
s_1^2 =	18.67	s_2^2 =	19.00	s_3^2 =	1.00	s_4^2 =	

Execute



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

10.3 Nonparametric Test for Location of Several Populations

10.3.1 Completely Randomized Design : 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.

- $h(n_1, n_2, \ldots, n_k)$: Kruskal-Wallis H distribution
- If there are tied values in the combined sample, assign the average of ranks.

10.3 Nonparametric Test for Location of Several Populations

10.3.1 Completely Randomized Design : Kruskal-Wallis Test

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 are tied values in the combined sample, assign the average of ranks.

Thank you