Introduction to Statistics and Data Science using *eStat*

Chapter 9 Testing Hypothesis for Several Population Means

9.1 Analysis of Variance for Experiments of Single Factor

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

9.1 Analysis of Variance for Experiments of Single Factor

- Examples to compare means of several populations.
  - 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?

- **Factor** is a variable used to distinguish populations, such as grade or rice.

9.1 Analysis of Variance for Experiments of Single Factor

[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 the data are in Table 9.1.1.

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

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

9.1 Analysis of Variance for Experiments of Single Factor

<Answer of Example 9.1.1>

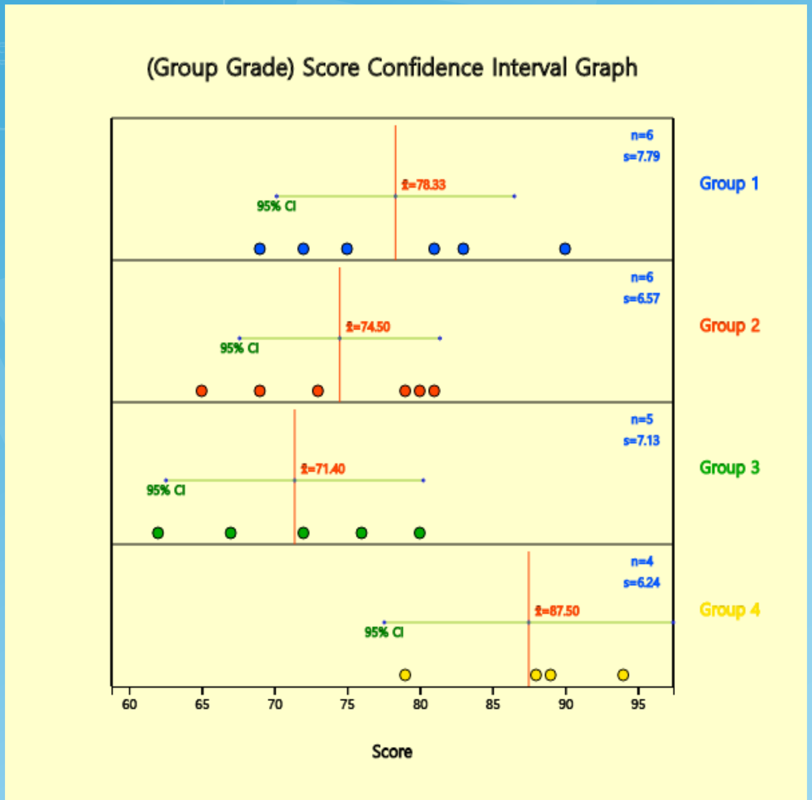
File: Ex911EnglishScoreByGrade.csv

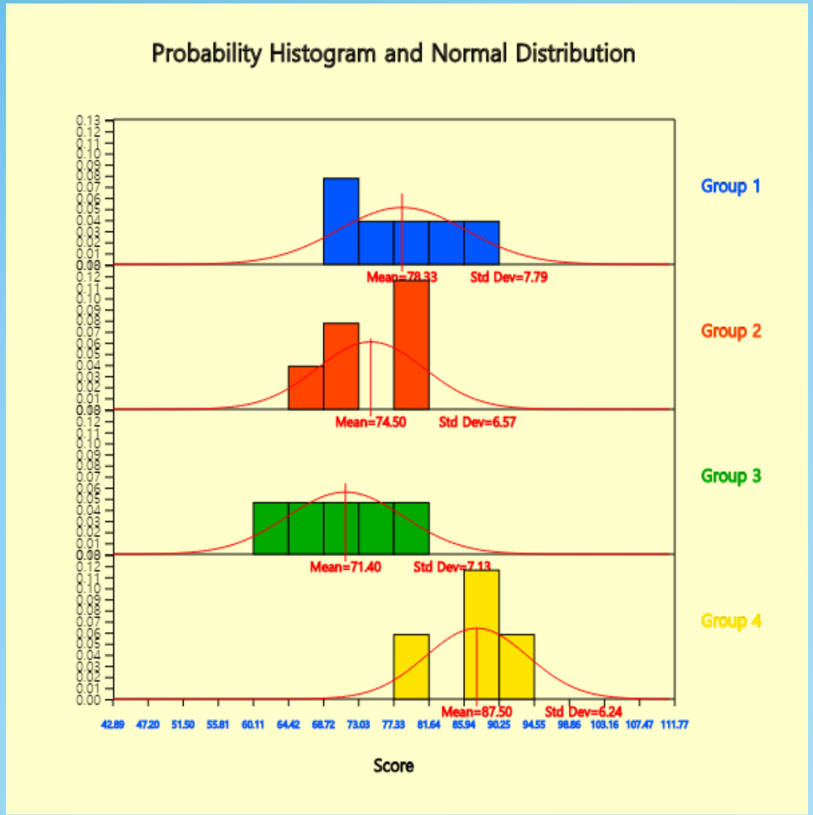
Analysis Var: 2: Score by Group: 1: Grade

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

SelectedVar: V2 by V1,

	Grade	Score	V3	V4
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		





Confidence Interval Graph | Histogram

$H_0 : \mu_1 = \mu_2 = \ldots = \mu_k$ H_1 : At least one pair of means is different

Significance Level α = ◉ 5% ○ 1% Confidence Level ◉ 95% ○ 99%

ANOVA F test | Standardized Residual Plot | Kruskal-Wallis Test

9.1 Analysis of Variance for Experiments of Single Factor

<Answer of Example 9.1.1>

2) Null hypothesis H_o : $\mu_1 = \mu_2 = \mu_3 = \mu_4$

Alternative hypothesis H_1 : *at least one pair of* μ_i *is not the same*

3) Between sum of squares (SSB) or Treatment sum of squares (SSTr)

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

⇨ If SSTr is close to zero, all sample means for four grades are similar.

Within sum of squares (SSW) or Error sum of squares (SSE)

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

9.1 Analysis of Variance for Experiments of Single Factor

<Answer of Example 9.1.1>

$$F_0 = \frac{\frac{SSTr}{(4-1)}}{\frac{SSE}{(21-4)}} = \frac{Treatment\ \ Mean\ Square\ (MSTr)}{Error\ Mean\ Square\ (MSE)} \quad \sim \quad F_{3,17}$$

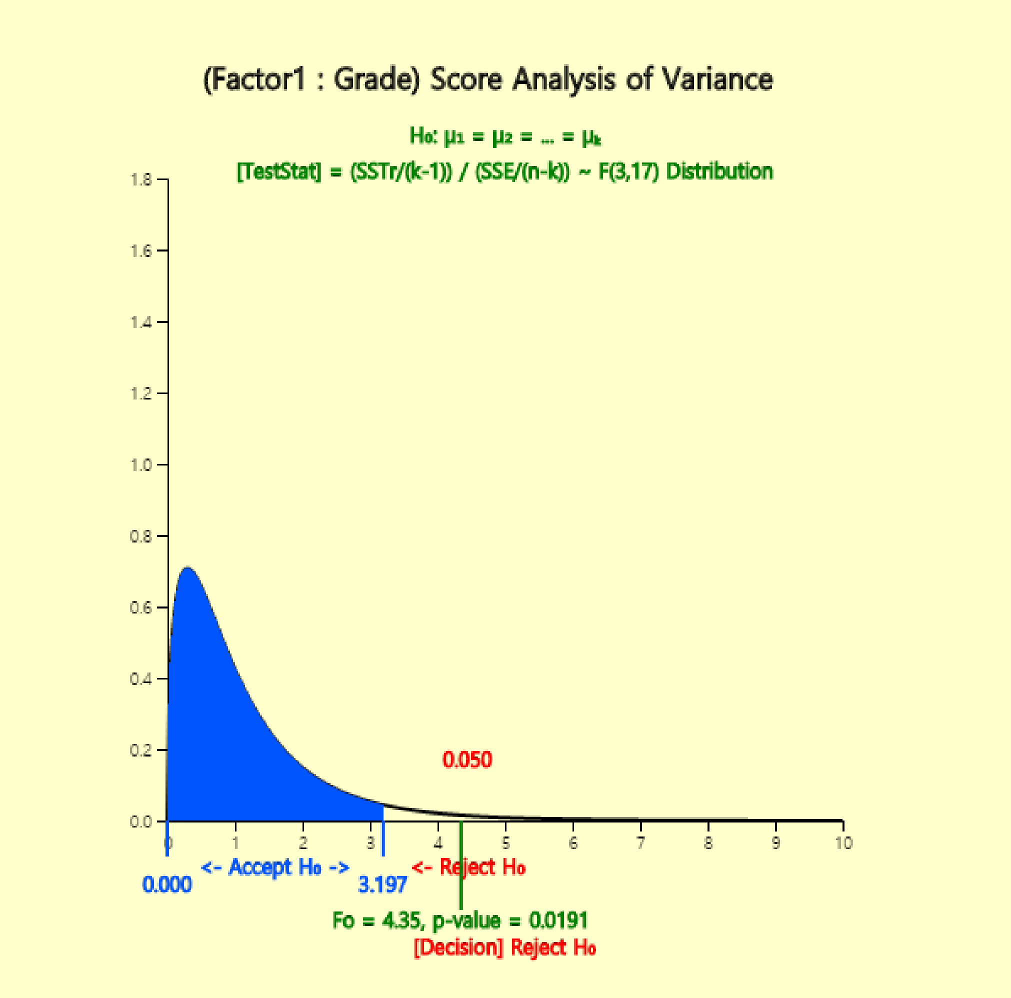
$$F_0 = \frac{\frac{643.633}{3}}{\frac{839.033}{17}} = 4.347 \qquad F_{3,17;0.05} = 3.20 \cdots$$

- **Hence Reject** $H_o : \mu_1 = \mu_2 = \mu_3 = \mu_4$
- **ANOVA Table**

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		

9.1 Analysis of Variance for Experiments of Single Factor

<Answer of Example 9.1.1>



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			

9.1 Analysis of Variance for Experiments of Single Factor

<Answer of Example 9.1.1>

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	81 75 69 90 72 83
Sample 2	65 80 73 79 81 69
Sample 3	72 67 62 76 80
Sample 4	89 94 79 88

[Sample Statistics]

$n_1 =$	6	$n_2 =$	6	$n_3 =$	5	$n_4 =$	4
$\bar{x}_1 =$	78.33	$\bar{x}_2 =$	74.50	$\bar{x}_3 =$	71.40	$\bar{x}_4 =$	87.50
$s_1^2 =$	60.67	$s_2^2 =$	43.10	$s_3^2 =$	50.80	$s_4^2 =$	39.00

Execute

9.1 Analysis of Variance for Experiments of Single Factor

Table 9.1.2 Notation of 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}	$\overline{Y}_{2\cdot}$
Level k	Y_{k1}	Y_{k2}	$\cdots$	Y_{kn_k}	$\overline{Y}_{k\cdot}$

- **ANOVA Model**

$$Y_{ij} = \mu_i + \varepsilon_{ij}$$
$$= \mu + \alpha_i + \varepsilon_{ij}, \quad i = 1, 2, \ldots, k \,;\, j = 1, 2, \ldots, n_i$$

- **Hypothesis**

H_0 : $\alpha_1 = \alpha_2 = \cdots = \alpha_k = 0$
H_1 : At least one pair of α_i is not equal to 0

9.1 Analysis of Variance for Experiments of Single Factor

Table 9.1.3 Analysis of variance table of one-way ANOVA

Factor	Sum of Squares	Degree of freedom	Mean Squares	F value
Treatment	SSTr	$k-1$	MSTr = SSTr / $(k-1)$	F_0 = MSTr/MSE
Error	SSE	$n-k$	MSE = SSE / $(n-k)$	
Total	SST	$n-1$	$(n=\sum_{i=1}^{k} n_i)$	

- **Total Sum of Squares** **SST** $= \sum_{i=1}^{k}\sum_{j=1}^{n_i}(Y_{ij} - \bar{Y}_{..})^2$
- **Treatment Sum of Squares** **SSTr** $= \sum_{i=1}^{k}\sum_{j=1}^{n_i}(\bar{Y}_{i.} - \bar{Y}_{..})^2$
- **Error Sum of Squares** **SSE** $= \sum_{i=1}^{k}\sum_{j=1}^{n_i}(Y_{ij} - \bar{Y}_{i.})^2$
- **SST = SSTr + SSE**
- **If $F_0 > F_{k-1,n-k\,;\,\alpha}$, then reject H_0**

Thank you