Introduction to Statistics and Data Science using *eStat*

Chapter 9 Testing Hypothesis for Several Population Means

9.1 ANOVA for Experiments of Single Factor

9.1.1 Multiple Comparison

9.1.2 Residual Analysis

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

9.1 Analysis of Variance for Experiments of Single Factor

9.1.1 Multiple Comparison

9.1.2 Residual Analysis

9.2 Design of Experiments for Sampling

9.2.1 Completely Randomized Design

9.2.2 Randomized block design

9.3 Analysis of Variance for Experiments of Two Factors

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

<Answer>

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

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

9.1 Analysis of Variance for Experiments of Single Factor

<Answer of Example 9.1.1>

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

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

- **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 Error	SSTr= 643.633 SSE = 839.033	4-1 21-4	MSTr = 643.633/3 MSE = 839.033/17	Fo = 4.347
Total	SST =1482.666	20		

9.1 Analysis of Variance for Experiments of Single Factor

9.1.1 Multiple Comparison

- **Hypothesis**

$$H_o : \mu_i = \mu_j$$
$$H_1 : \mu_i \neq \mu_j \qquad i = 1, 2, \dots, k-1 ; \quad j = i+1, i+2, \dots, k$$

- **Tukey's Honestly Significant Difference (HSD) Test**

If $|\bar{y}_i - \bar{y}_j| > HSD_{ij}$, then Reject H_o

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

$q_{k,n-k;\,\alpha}$ is the studentized range distribution

9.1 Analysis of Variance for Experiments of Single Factor

9.1.1 Multiple Comparison

HSD Studentized Range Dist.

Menu

α = 5% 1%

Percentile Table Table Save

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

9.1 Analysis of Variance for Experiments of Single Factor

[Example 9.1.2] In [Example 9.1.1], the analysis variance of English scores by grade concluded that the null hypothesis was rejected and the average English scores for each grade were not all the same.

- Apply multiple comparisons to check where the differences exist among each school grade with a significant level of 5%.
- Use『eStat』to check the results.

<Answer>

- Hypothesis $H_o: \mu_i = \mu_j, \; H_1: \mu_i \neq \mu_j \quad i = 1, 2, \ldots, 3; \; j = i+1, i+2, \ldots, 4$
- HSD Test

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

9.1 Analysis of Variance for Experiments of Single Factor

<Answer of Example 9.1.2>

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

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

$$\mathrm{HSD}_{12} = q_{k,n-k;0.05} \cdot \sqrt{\frac{1}{2}\left(\frac{1}{n_1}+\frac{1}{n_2}\right)\mathrm{MSE}} = q_{4,21-4;0.05} \cdot \sqrt{\frac{1}{2}\left(\frac{1}{6}+\frac{1}{6}\right)49.355} = 11.530$$

Therefore, accept H_0.

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

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

$$\mathrm{HSD}_{13} = q_{k,n-k;0.05} \cdot \sqrt{\frac{1}{2}\left(\frac{1}{n_1}+\frac{1}{n_3}\right)\mathrm{MSE}} = q_{4,21-4;0.05} \cdot \sqrt{\frac{1}{2}\left(\frac{1}{6}+\frac{1}{5}\right)49.355} = 12.092$$

Therefore accept H_0.

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

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

$$\mathrm{HSD}_{14} = q_{k,n-k;0.05} \cdot \sqrt{\frac{1}{2}\left(\frac{1}{n_1}+\frac{1}{n_4}\right)\mathrm{MSE}} = q_{4,21-4;0.05} \cdot \sqrt{\frac{1}{2}\left(\frac{1}{6}+\frac{1}{4}\right)49.355} = 12.891$$

Therefore accept H_0.

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

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

$$\mathrm{HSD}_{23} = q_{k,n-k;0.05} \cdot \sqrt{\frac{1}{2}\left(\frac{1}{n_2}+\frac{1}{n_3}\right)\mathrm{MSE}} = q_{4,21-4;0.05} \cdot \sqrt{\frac{1}{2}\left(\frac{1}{6}+\frac{1}{5}\right)49.355} = 12.092$$

Therefore accept H_0.

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

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

$$\mathrm{HSD}_{24} = q_{k,n-k;0.05} \cdot \sqrt{\frac{1}{2}\left(\frac{1}{n_2}+\frac{1}{n_4}\right)\mathrm{MSE}} = q_{4,21-4;0.05} \cdot \sqrt{\frac{1}{2}\left(\frac{1}{6}+\frac{1}{4}\right)49.355} = 12.891$$

Therefore, reject H_0.

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

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

$$\mathrm{HSD}_{34} = q_{k,n-k;0.05} \cdot \sqrt{\frac{1}{2}\left(\frac{1}{n_3}+\frac{1}{n_4}\right)\mathrm{MSE}} = q_{4,21-4;0.05} \cdot \sqrt{\frac{1}{2}\left(\frac{1}{5}+\frac{1}{4}\right)49.355} = 13.396$$

Therefore, reject H_0.

9.1 Analysis of Variance for Experiments of Single Factor

<Answer of Example 9.1.2>

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

Testing Means * 95%, ** 99%	3 (Group 3) 71.40	2 (Group 2) 74.50	1 (Group 1) 78.33	4 (Group 4) 87.50
3 (Group 3) 71.40				*
2 (Group 2) 74.50				*
1 (Group 1) 78.33				
4 (Group 4) 87.50	*	*		

9.1 Analysis of Variance for Experiments of Single Factor

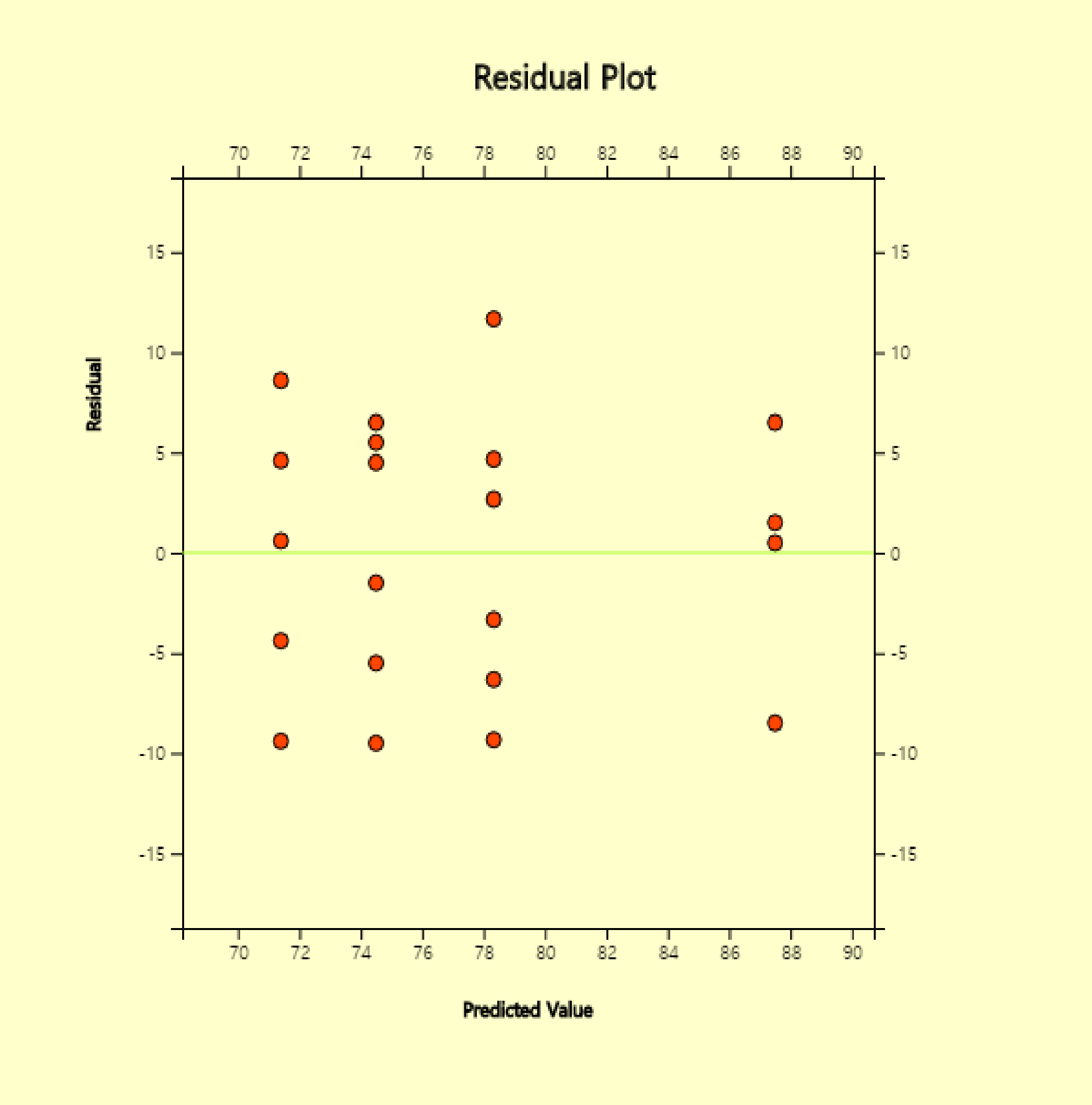
9.1.2 Residual Analysis

- **ANOVA test is based on assumptions about the error term ϵ_{ij}.**
  - ⇨ ϵ_{ij}'s are independent of each other (independence)
  - ⇨ each variance of ϵ_{ij} is constant , σ^2 (homoscedasticity)
  - ⇨ each ϵ_{ij} is normally distributed (normality)

- **Validity of these assumptions should always be investigated.**
  - ⇨ ϵ_{ij} can not be observed
  - ⇨ residual, $Y_{ij} - \overline{Y}_{i\cdot}$, as estimate of ϵ_{ij} is used to check assumptions.
  - ⇨ residual analysis

9.1 Analysis of Variance for Experiments of Single Factor

9.1.2 Residual Analysis



Thank you