4. POST HOC ANALYSES

In real one-way ANOVA, The usual practice is to perform the ***overall F test*** first on the overall hypothesis H_0: all group means are *equal*, vs. H_1: at least two group means *differ*.

Under aforesaid one-way ANOVA model, $F \sim F_{k-1,n-k}$ if the overall null hypothesis is true.

Reject H_0 at nominal level $\alpha \in (0,1)$ if $F > F_{k-1,n-k,1-\alpha}$ (namely, $p < \alpha$).

If the overall null hypothesis is rejected, then specific groups are compared to determine which of the groups have different means.

The most common post hoc tests are:

- Bonferroni Procedure
- Duncan's new multiple range test (MRT)
- Dunn's Multiple Comparison Test
- Fisher's Least Significant Difference (LSD)
- Holm-Bonferroni Procedure
- Newman-Keuls
- Rodger's Method
- Scheffé's Method
- Tukey's Test (see also: Studentized Range Distribution)
- Dunnett's correction
- Benjamini-Hochberg (BH) procedure

(1) FISHER'S LEAST SIGNIFICANT DIFFERENCE (LSD) METHOD

The most sensitive method

Consider *two* ***independent*** *samples*, $\{y_{i1}: i = 1, \ldots, n_1\}$ iid $\sim N(\mu_1, \sigma^2)$, and $\{y_{j2}: j = 1, \ldots, n_2\} \sim N(\mu_2, \sigma^2)$.

We can perform ***double-sided t test*** of hypothesis $H_0: \mu_1 = \mu_2$ vs. $H_a: \mu_1 \neq \mu_2$, where the t statistic is:

$$t = \frac{\bar{y}_1 - \bar{y}_2}{s_c \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}.$$

We can estimate the common variance σ^2 by the pooled sample variance

$$s_C^2 = \frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{n_1 + n_2 - 2},$$

T-test

In one-way ANOVA of k groups, we have $\boldsymbol{s_p} = \sqrt{\mathbf{MSW}}$ since we have k groups and assume *homoscedasticity*.

$$s_{\mathrm{p}}^2 = \frac{\sum_{i=1}^{k}(n_i - 1)s_i^2}{\sum_{i=1}^{k}(n_i - 1)} = \frac{\sum_{i=1}^{k}(n_i - 1)s_i^2}{n - k} = MSW$$

PAIRWISE *T* TEST

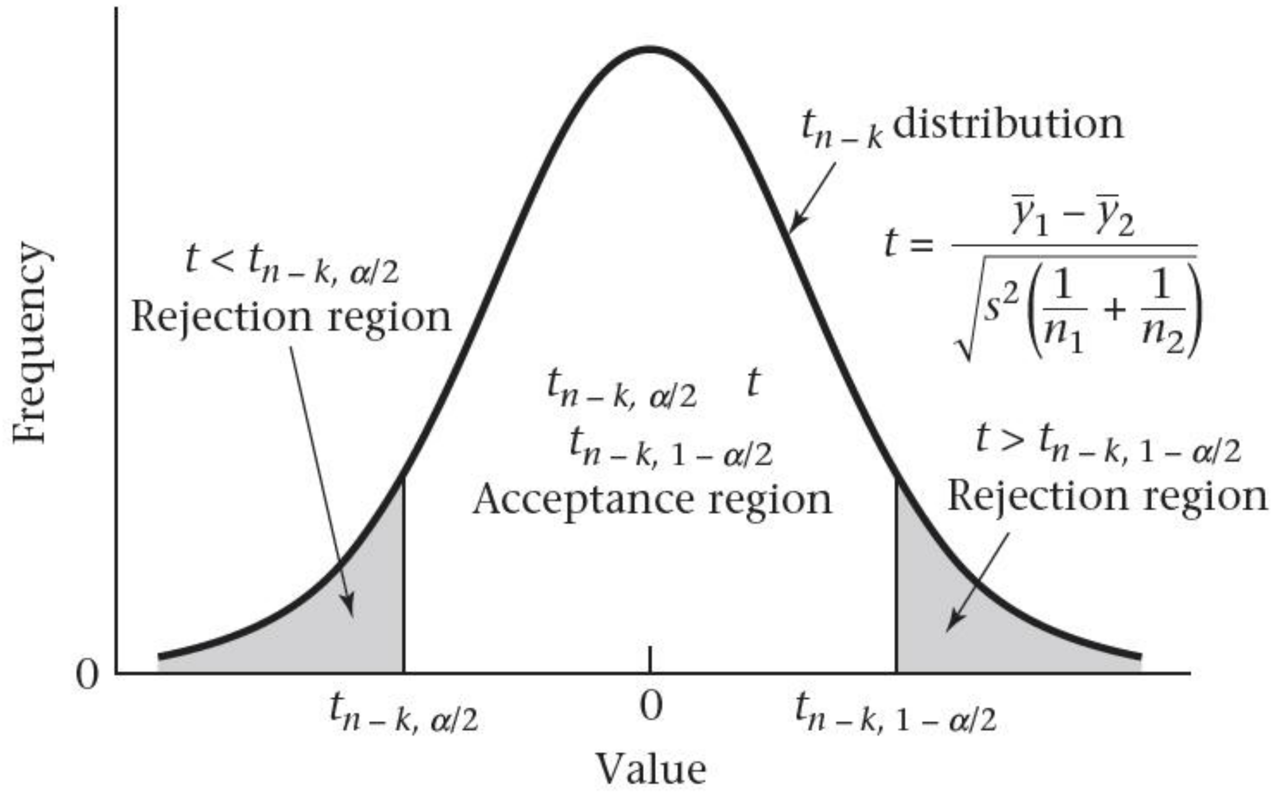

Figure 4.1. Acceptance and rejection regions for the *t* test for the comparison of pairs of groups in one-way ANOVA (LSD approach)

Note 1: As the estimators of σ^2, the pooled sample variance $s_p^2 = \sqrt{\mathrm{WMS}}$ based on all k groups will be more accurate than that obtained from using **any two groups** because s_p^2 is based on more information.

This is the principal advantage of performing the t tests in the framework of a one-way ANOVA rather than by considering each pair of groups separately and performing t tests for two independent samples.

Note 2: The LSD procedure should only be used when ANOVA leads to significant differences and only a few pre-planned comparisons are made.

For k groups, the number of all pairwise comparisons is $m = k(k-1)/2$. We commit Type error when all the k means are identical but we reject the null hypothesis H_{0j} of any pairwise comparisons. The overall (**experiment-wise**) type I error rate is given by

$$\alpha_{LSD} = \Pr\left(\bigcup_{j=1}^{m}\{\text{reject } H_{0j}\} \mid \text{all } H_{0j} \text{ are true}\right).$$

In general, LSD does not control for *experiment-wise error rate* at *pairwise rate* α (**Figure 4.2**).

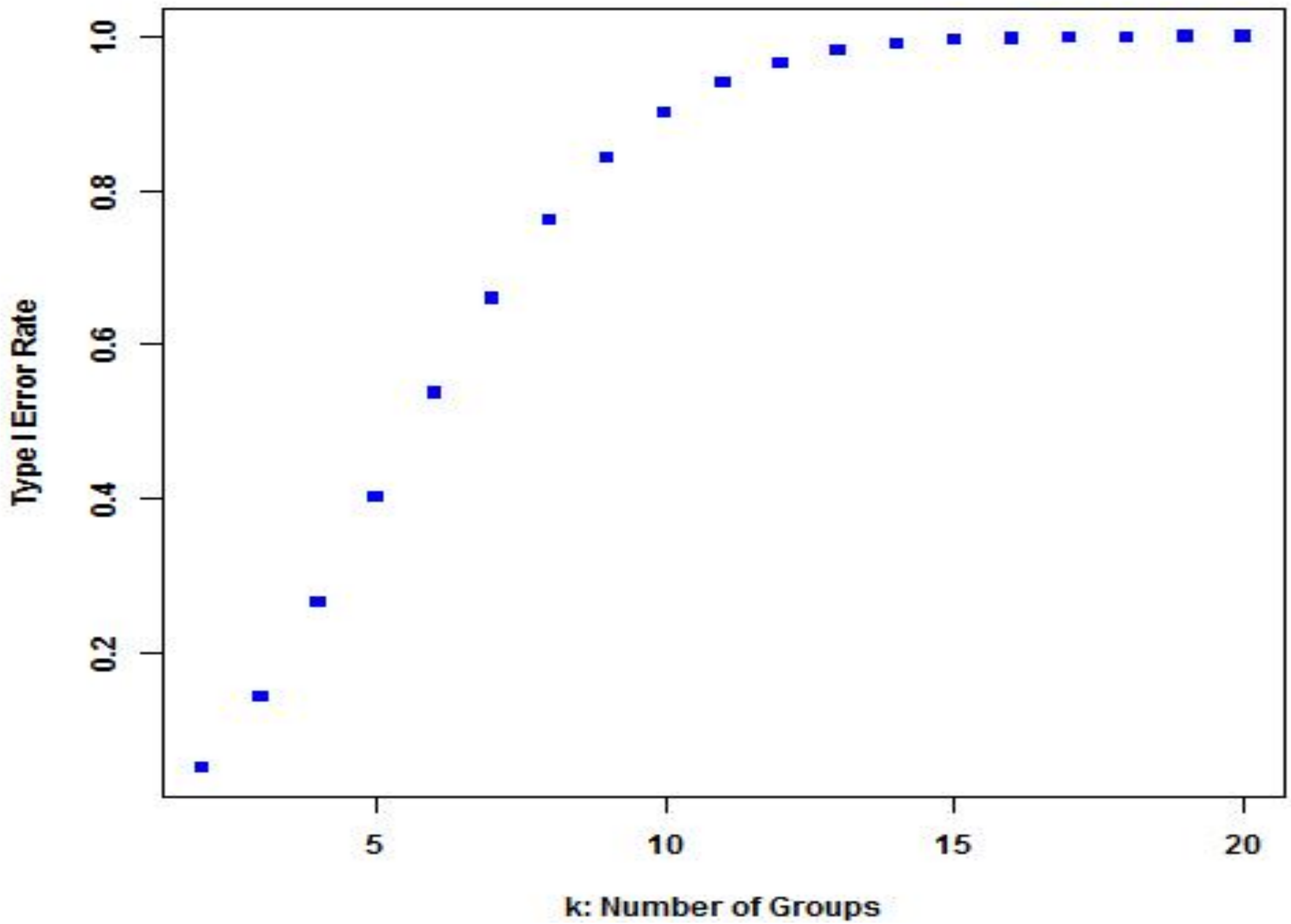

Figure 4.2: *Experiment-wise Type I error rate* of LSD procedure. For each pairwise comparison, the pairwise nominal level is $\alpha = 0.05$.

EXAMPLE 4.1 (NUTRITION):

Researchers compared *protein intake* among *three groups* of postmenopausal women:

(1) women eating a standard American diet (STD),

(2) women eating a lactoovo-vegetarian diet (LAC), and

(3) women eating a strict vegetarian diet (VEG).

The *mean* and *sd* for protein intake (mg) is presented in **Table 4.1**.

Table 4.1: Protein intake (mg) among three dietary groups of postmenopausal women			
Group	$\bar{y}_i$	s_i	n_i
STD	75	9	10
LAC	57	13	10
VEG	47	17	6

Problems

1) Perform an overall F test to compare the means of the three groups.

2) Compare the means of each specific pair of groups using the LSD method.

SOLUTION 1:

The number of groups are $k = 3$, groups sizes are $(n_1, n_2, n_3) = (10,10,6)$ and total sample size is $n = 26$.

(1) Using the *short form equations*, we obtain:

$$\text{SSB} = \sum_{i=1}^{k} n_i(\bar{y}_i)^2 - \frac{(y_{..})^2}{n}$$

$$= [10(75)^2 + 10(57)^2 + 6(47)^2] - \frac{[10\times75+10\times57+6\times47]^2}{26}$$

$$= 101994 - 98707.85$$

$$= 3286.154,$$

$$\text{SSW} = \sum_{i=1}^{k} (n_i - 1)s_i^2$$
$$= 9(9)^2 + 9\,(13)^2 + 5(17)^2$$
$$= 729 + 1521 + 1445 = 3695.$$

It follows that

$$\text{MSB} = \frac{\mathbf{3286.154}}{\mathbf{2}} = 1643.077,$$

$$\text{MSW} = \frac{3695}{26 - 3} = \frac{3695}{23} = 160.6522,$$

$$\text{F} = \frac{\text{MSB}}{\text{MSW}} = \frac{1643.077}{160.6522} = 10.22754.$$

Table 4.2: ANOVA for overall F test of nutrition.					
Source	DF	Sum of Squares	Mean Square	F statistic	P value
Between	2	3286.345	1643.077	10.22754	0.000664
Within	23	3695	160.6522		
Total	25	7521.154			

Solution 2:

(2) Pairwise LSD method (t test)

The pooled sample variance is $s_p^2 = 160.6522$ and the square root is $s_p = 12.67498$. For pairwise level $\alpha = 0.05$, the critical value is $t_{n-k,1-\alpha/2} = t_{23,0.975} = 2.068658$. Hence, $\tau \stackrel{\text{def}}{=} t_{23,0.975} \times s_p = (2.068658)(12.67498) = 26.22019878$. The LSDs are given below:

$$\mathrm{LSD}(1,2) = \tau \times \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}$$

$$= 26.22019878 \times \sqrt{\frac{1}{10} + \frac{1}{10}} = 11.71,$$

$$\mathrm{LSD}(1,3) = \tau \times \sqrt{\frac{1}{n_1} + \frac{1}{n_3}}$$

$$= 26.22019878 \times \sqrt{\frac{1}{10} + \frac{1}{6}}$$

$$= 13.52,$$

$\mathrm{LSD}(2,3) = \mathrm{LSD}(1,3) = 13.52$ [because $n_1 = n_3$]

The *t* statistics are given below:

$$t(1,2) = \frac{\bar{y}_1 - \bar{y}_2}{s_p \times \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} = \frac{18}{12.67498 \times \sqrt{\frac{1}{10} + \frac{1}{10}}}$$

$$= \frac{18}{12.67498 \times 0.4472136} = 3.175486,$$

$$t(1,3) = \frac{\bar{y}_1 - \bar{y}_3}{s_p \times \sqrt{\frac{1}{n_1} + \frac{1}{n_3}}} = \frac{28}{12.67498 \times \sqrt{\frac{1}{10} + \frac{1}{6}}}$$

$$= \frac{28}{12.67498 \times 0.5163978} = 4.277858,$$

$$t(2,3) = \frac{\bar{y}_2 - \bar{y}_3}{s_p \times \sqrt{\frac{1}{n_2} + \frac{1}{n_3}}} = \frac{10}{12.67498 \times \sqrt{\frac{1}{10} + \frac{1}{6}}}$$

$$= \frac{10}{12.67498 \times 0.5163978} = 1.527806.$$

The results are summarized in **Table 4.3 (*: significant at pairwise level 0.05; **: not significant at pairwise level 0.05)**. There are *highly significant evidence* to conclude that the STD women differ from the LAC and the VEG women. There are no significant differences (?) between the LAC and the VEG women.

Table 4.3: Comparisons of specific pairs of groups for the nutrition data using the LSD t test approach

Groups compared	diff	LSD	t statistic	P value	Decision
STD, LAC	18	11.71	3.175486	0.004218	*
STD, VEG	28	13.52	4.277858	0.000282	*
LAC, VEG	10	13.52	1.527806	0.140197	**

(2) BONFERRONI CORRECTION FOR THE LSD

For k groups, the number of all pairwise comparisons is $m = C_k^2 = k(k-1)/2$.

$$a^{*} = \frac{a}{c_k^2} = \frac{2\alpha}{k(k-1)}$$

Comparison of Pairs of Groups in One-Way ANOVA—Bonferroni Multiple-Comparisons Procedure

Suppose we wish to compare two specific groups, arbitrarily labeled as group 1 and group 2, among k groups. To test the hypothesis $H_0\colon \alpha_1 = \alpha_2$ vs. $H_1\colon \alpha_1 \neq \alpha_2$, use the following procedure:

(1) Compute the pooled estimate of the variance s^2 = Within MS from the one-way ANOVA.

(2) Compute the test statistic

$$t = \frac{\bar{y}_1 - \bar{y}_2}{\sqrt{s^2\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}}$$

(3) For a two-sided level α test, let $\boxed{\alpha^* = \alpha \Big/ \binom{k}{2}}$

If $\quad t > t_{n-k,1-\alpha^*/2} \quad$ or $\quad t < t_{n-k,\alpha^*/2} \quad$ then reject H_0

If $\quad t_{n-k,\alpha^*/2} \le t \le t_{n-k,1-\alpha^*/2} \quad$ then accept H_0

Rational

- For k groups, the number of all pairwise comparisons is $m = C_k^2 = k(k-1)/2$.
- Let E be the event that at least one of the two-group comparisons is statistically significant ($reject\ H_0$), $\Pr(E)$ is sometimes referred to as the "**pairwise** type I error rates".
- We wish to find an α*

$$\alpha^* = \Pr\left(\bigcup_{j=1}^{m} \{\text{reject } H_{0j}\} \mid \text{all } H_{0j} \text{ are true}\right)$$

$$\leq \sum_{j}^{m} \Pr\left(\text{reject } H_{0j} \mid H_{0j}\ is \text{ true}\right) = \alpha.$$

Rational

- $\Pr(\bar{E}) = \Pr($none of the two-group comparisons is statistically significant$) = 1-\alpha$
- If each of the two-group comparisons were independent, then from the multiplication law of probability, $\Pr(\bar{E}) = (1-\alpha^*)^m$
- Therefore, $1-\alpha = (1-\alpha^*)^m$
- If α^* is small, then it can be shown that $(1-\alpha^*)^m$ can be approximated by $1 - m\alpha^*$
- $1-\alpha \cong (1-\alpha^*)^m$ or $\alpha^* \cong \frac{a}{c_k^2} = \frac{2\alpha}{k(k-1)}$

Bonferroni Correction reduces the risk of committing *Type I Error*, but may increase the risk of committing Type II error (**reducing statistical power**). **Tukey's HSD method** is a frequently used alternative.

(3)TUKEY'S HONESTLY SIGNIFICANT DIFFERENCE (HSD) METHOD

For an arbitrary group (i_1, i_2), reject the pairwise null of $\mu_{i_1} = \mu_{i_2}$ if $|\bar{y}_{i_1} - \bar{y}_{i_2}| > \text{HSD}(i_1, i_2)$, where $\text{HSD}(i_1, i_2)$ is determined by k, overall level α and group size. This method control overall type I error rate.

Equal sample sizes

Let $n_1 = \ldots = n_k$. Define

$$\text{HSD} = \frac{s_p \times q_{k,n-k,1-\alpha}}{\sqrt{n_1}},$$

where $q_{k,n-k,1-\alpha}$ is the $(1-\alpha)100$ percentile of the Studentized Range:

$$q = \frac{\sqrt{n_1}(\max\{\bar{y}_1, \ldots, \bar{y}_k\} - \min\{\bar{y}_1, \ldots, \bar{y}_k\})}{s_p}.$$

See the **Appendix Table** for $\alpha = 0.05$.

APPENDIX: PERCENTAGE POINTS OF THE STUDENTIZED RANGE FOR 2 THROUGH 20 TREATMENTS UPPER 5% POINTS

Error df	2	3	4	5	6	7	8	9	10
1	17.97	26.98	32.82	37.08	40.41	43.12	45.40	47.36	49.07
2	6.08	8.33	9.80	10.88	11.74	12.44	13.03	13.54	13.99
3	4.50	5.91	6.82	7.50	8.04	8.48	8.85	9.18	9.46
4	3.93	5.04	5.76	6.29	6.71	7.05	7.35	7.60	7.83
5	3.64	4.60	5.22	5.67	6.03	6.33	6.58	6.80	6.99
6	3.46	4.34	4.90	5.30	5.63	5.90	6.12	6.32	6.49
7	3.34	4.16	4.68	5.06	5.36	5.61	5.82	6.00	6.16
8	3.26	4.04	4.53	4.89	5.17	5.40	5.60	5.77	5.92
9	3.20	3.95	4.41	4.76	5.02	5.24	5.43	5.59	5.74
10	3.15	3.88	4.33	4.65	4.91	5.12	5.30	5.46	5.60
11	3.11	3.82	4.26	4.57	4.82	5.03	5.20	5.35	5.49
12	3.08	3.77	4.20	4.51	4.75	4.95	5.12	5.27	5.39
13	3.06	3.73	4.15	4.45	4.69	4.88	5.05	5.19	5.32
14	3.03	3.70	4.11	4.41	4.64	4.83	4.99	5.13	5.25
15	3.01	3.67	4.08	4.37	4.59	4.78	4.94	5.08	5.20
16	3.00	3.65	4.05	4.33	4.56	4.74	4.90	5.03	5.15
17	2.98	3.63	4.02	4.30	4.52	4.70	4.86	4.99	5.11
18	2.97	3.61	4.00	4.28	4.49	4.67	4.82	4.96	5.07
19	2.96	3.59	3.98	4.25	4.47	4.65	4.79	4.92	5.04
20	2.95	3.58	3.96	4.23	4.45	4.62	4.77	4.90	5.01
24	2.92	3.53	3.90	4.17	4.37	4.54	4.68	4.81	4.92
30	2.89	3.49	3.85	4.10	4.30	4.46	4.60	4.72	4.82
40	2.86	3.44	3.79	4.04	4.23	4.39	4.52	4.63	4.73
60	2.83	3.40	3.74	3.98	4.16	4.31	4.44	4.55	4.65
120	2.80	3.36	3.68	3.92	4.10	4.24	4.36	4.47	4.56
∞	2.77	3.31	3.63	3.86	4.03	4.17	4.29	4.39	4.47

Unequal sample sizes

Tukey-Kramer method:

$$\mathrm{HSD(i,j)} = \frac{s_p \times q_{k,n-k,1-\alpha}}{\sqrt{2}} \sqrt{\frac{1}{n_i} + \frac{1}{n_j}},$$

where ***n*** is the total sample size.

Harmonic Mean method:

$$\mathrm{HSD} = \frac{s_p \times q_{k,n-k,1-\alpha}}{\sqrt{n_h}},$$

where $n_h = \left(\frac{1}{k}\sum_{i=1}^{k} \frac{1}{n_i}\right)^{-1}$ is *Harmonic Mean* of the sample sizes.

Note 3: The *Tukey-Kramer* method results in smaller Type I error rate than the stated α more often (namely, *less conservative*) than the harmonic mean method.

Using the data in Table 4.1, *perform Tukey-Kramer* and *Harmonic Mean* methods at overall level $\alpha = 0.05$ to identify which specific underlying means are different.

Solution:

Since the sample sizes are unequal and the overall level is $\alpha = 0.05$, the $(1-\alpha)100$ percentile of the Studentized Range is $q_{k,n-k,1-\alpha} = q_{3,23,0.95} \approx 3.58$.

Tukey-Kramer : The HSDs are given below:

$$\text{HSD}(1,2) = \frac{s_p \times q_{k,n-k,1-\alpha}}{\sqrt{2}}\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}$$

$$= \frac{12.67498 \times 3.58}{\sqrt{2}}\sqrt{\frac{1}{10}+\frac{1}{10}} = 14.34929,$$

$$\text{HSD}(1,3) = \frac{s_p \times q_{k,n-k,1-\alpha}}{\sqrt{2}}\sqrt{\frac{1}{n_1}+\frac{1}{n_3}}$$

$$= \frac{12.67498 \times 3.58}{\sqrt{2}}\sqrt{\frac{1}{10}+\frac{1}{6}} = 19.42903,$$

and $\text{HSD}(2,3) = \text{HSD}(1,3) = 19.42903.$

The comparison results are summarized in **Table 4.4 (*Significant at overall level 0.05. **Not significant at overall level 0.05).** Conclusion is the same as that in ***Example 4.1***.

Table 4.4: Comparisons of specific pairs of groups for the nutrition data using the Tukey-Kramer method			
Groups compared	difference	HSD	Decision
STD, LAC	18	14.34	Significant*
STD, VEG	28	19.43	Significant*
LAC, VEG	10	19.43	Not significant**

(4) *Harmonic Mean method*

The Harmonic Mean of the sample sizes is:

$$n_h = \left(\frac{1}{k}\sum_{i=1}^{k}\frac{1}{n_i}\right)^{-1} = \left(\frac{\frac{1}{10}+\frac{1}{10}+\frac{1}{6}}{3}\right)^{-1} = 8.18.$$

It follows that

$$\mathrm{HSD} = \frac{s_p \times q_{k,n-k,1-\alpha}}{\sqrt{n_h}} = \frac{12.67498 \times 3.58}{\sqrt{8.18}} = 15.8655.$$

The comparison results are summarized in **Table 4.5 (*Significant at overall level 0.05. **Not significant at overall level 0.05).** Conclusion is the same as that in ***Example 14.1.***

Table 4.5: Comparisons of specific group pairs for the nutrition data using the Harmonic Mean method			
Groups compared	difference	HSD	Decision
STD, LAC	18	15.8655	Significant*
STD, VEG	28	15.8655	Significant*
LAC, VEG	10	15.8655	Not significant**