Multisample Inference 12

Solved Problems

12.1 Introduction to the One-Way Analysis of Variance

In Chapter 8 we were concerned with comparing the means of two normal distributions using the two-sample t test for independent samples. Frequently, the means of more than two distributions need to be compared. Therefore, the t test methodology generalizes nicely in this case to a procedure called the *one-way analysis of variance* (*ANOVA*). Question: How can the means of these k groups can be compared?

12.2 One-Way ANOVA-Fixed-Effects Model

Suppose there are k groups with n_i observations in the i^{th} group. The j^{th} observation in the i^{th} group will be denoted by y_{ij}. The typical data used for constructing a one-way ANOVA table would appear as shown in the Table below:

Group	Observations				Totals	Averages
1	y_{11}	y_{12}	⋯	y_{1n_1}	$y_{1.}$	$\bar{y}_{1.}$
2	y_{21}	y_{22}	⋯	y_{2n_2}	$y_{2.}$	$\bar{y}_{2.}$
⋮	⋮	⋮	⋮⋮⋮	⋮	⋮	⋮
k	y_{k1}	y_{k2}	⋯	y_{kn_k}	$y_{k.}$	$\bar{y}_{k.}$
		Total			$y_{..}$	

Let's assume the following model (*Fixed-Effects Model*):

EQUATION 12.1

$$y_{ij} = \mu + \alpha_i + e_{ij}$$

where μ is a constant, α_i is a constant specific to the i^{th} group, and e_{ij} is an error term, which is normally distributed with mean 0 and variance σ^2. Thus, a typical observation from the i^{th} group is normally distributed with mean $\mu + \alpha_i$ and variance σ^2. The parameters in Equation 12.1 can be interpreted as follows:

EQUATION 12.2

Interpretation of the Parameters of a One-Way ANOVA Fixed-Effects Model

(1) μ represents the underlying mean of all groups taken together.

(2) α_i represents the difference between the mean of the *i*th group and the overall mean.

(3) e_{ij} represents random error about the mean $\mu + \alpha_i$ for an individual observation from the *i*th group.

Some Notations

- It is not possible to estimate both the overall constant μ as well as the k constants α_i, which are specific to each group. The reason is that we only have k observed mean values for the k groups, which are used to estimate $k + 1$ parameters.
- We need to constrain the parameters so that only k parameters will be estimated.
- Some typical constraints are:
  (1) the sum of the α_i's is set to 0, or
  (2) the α_i for the last group (αk) is set to 0.

In this text, we will use the former approach in our Fixed-Effects Model and the Group Means are compared within the context of this model.

DEFINITION 12.1

The model in Equation 12.1 is a **one-way analysis of variance**, or a **one-way ANOVA model**. With this model, the means of an arbitrary number of groups, each of which follows a normal distribution with the same variance, can be compared. Whether the variability in the data comes mostly from variability within groups or can truly be attributed to variability between groups can also be determined.

12.3 Hypothesis Testing in One-Way ANOVA-Fixed-Effects Model

The null hypothesis (H_0) and the alternative hypothesis (H_1) can be stated as follows:

- The null hypothesis (H_0) in this case is that *the underlying mean of each of the k groups is the same*. This hypothesis is equivalent to stating that each $\alpha_i = 0$ because the α_i sum up to 0 (*that is*: H_0: all $\alpha_i = 0$).

- The alternative hypothesis (H_1) is that *at least two of the group means are not the same*. This hypothesis is equivalent to stating that at least one $\alpha_i \neq 0$ (*that is*: H_1: at least one $\alpha_i \neq 0$).

Thus, we wish to test the hypothesis: H_0: all $\alpha_i = 0$ *versus* H_1: at least one $\alpha_i \neq 0$

12.3.1 F Test for Overall Comparison of Group Means

The mean for the i^{th} group will be denoted by $\bar{y}_i$, and the mean over all groups by $\bar{\bar{y}}$. The deviation of an individual observation (y_{ij}) from the overall mean ($\bar{\bar{y}}$), that is, $(y_{ij} - \bar{\bar{y}})$, can be represented as follows:

EQUATION 12.3

$$y_{ij} - \bar{\bar{y}} = (y_{ij} - \bar{y}_i) + (\bar{y}_i - \bar{\bar{y}})$$

where:

(i) $(y_{ij} - \bar{y}_i)$: represents the deviation of an individual observation (y_{ij}) from the group mean ($\bar{y}_i$) for that observation and is an indication of *within-group variability*.

(ii) $(\bar{y}_i - \bar{\bar{y}})$: represents the deviation of a group mean ($\bar{y}_i$) from the overall mean ($\bar{\bar{y}}$) and is an indication of *between-group variability*.

Now, if both sides of Equation 12.3 are squared and the squared deviations are summed over all observations over all groups, then the following relationship is obtained:

EQUATION 12.4

$$\sum_{i=1}^{k}\sum_{j=1}^{n_i}\left(y_{ij} - \bar{\bar{y}}\right)^2 = \sum_{i=1}^{k}\sum_{j=1}^{n_i}\left(y_{ij} - \bar{y}_i\right)^2 + \sum_{i=1}^{k}\sum_{j=1}^{n_i}\left(\bar{y}_i - \bar{\bar{y}}\right)^2$$

Thus, this will leads to three important definitions given as follows:

DEFINITION 12.2 The term

$$\sum_{i=1}^{k}\sum_{j=1}^{n_i}\left(y_{ij}-\bar{\bar{y}}\right)^2$$

is called the **Total Sum of Squares (Total SS).**

DEFINITION 12.3 The term

$$\sum_{i=1}^{k}\sum_{j=1}^{n_i}\left(y_{ij}-\bar{y}_i\right)^2$$

is called the **Within Sum of Squares (Within SS).**

DEFINITION 12.4 The term

$$\sum_{i=1}^{k}\sum_{j=1}^{n_i}\left(\bar{y}_i-\bar{\bar{y}}\right)^2$$

is called the **Between Sum of Squares (Between SS).**

Thus, the relationship in Equation 12.4 can be written as follows:

$$Total\ SS = Within\ SS + Between\ SS$$
$$SST = SSW + SSB$$

To perform the hypothesis test, it is easier to use the short computational form for the Within SS and Between SS in Equation 12.5 as follows:

EQUATION 12.5

Short Computational Form for the Between SS and Within SS

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

$$\text{Within SS} = \sum_{i=1}^{k}(n_i-1)s_i^2$$

where $y_{..}$ = sum of the observations across all groups, n = total number of observations over all groups, and s_i^2 = sample variance for the ith group.

Finally, the following two definitions are also important:

DEFINITION 12.5 **Between Mean Square = Between MS** = Between SS/$(k - 1)$

DEFINITION 12.6 **Within Mean Square = Within MS** = Within SS/$(n - k)$

The significance test will be based on the ratio of the **Between MS** to the **Within MS**. Then:

- If the ratio is large, then we reject H_0.
- If the ratio is small, we accept (or *fail to reject*) H_0.

Notation

Under H_0, the ratio of **Between MS** to **Within MS** follows an **F-distribution** with degrees of freedom $(k - 1)$ and $(n - k)$. Thus, the following **test procedure** for a significance level α test is used.

EQUATION 12.6

Overall *F* Test for One-Way ANOVA

To test the hypothesis H_0: $\alpha_i = 0$ for all i vs. H_1: at least one $\alpha_i \neq 0$, use the following procedure:

(1) Compute the Between SS, Between MS, Within SS, and Within MS using Equation 12.5 and Definitions 12.5 and 12.6.

(2) Compute the test statistic F = Between MS/Within MS, which follows an F distribution with $k - 1$ and $n - k$ *df* under H_0.

(3) If $F > F_{k-1,n-k,1-\alpha}$ then reject H_0
If $F \le F_{k-1,n-k,1-\alpha}$ then accept H_0

(4) The exact p-value is given by the area to the right of F under an $F_{k-1,n-k}$ distribution $= Pr(F_{k-1,n-k} > F)$.

The acceptance and rejection regions for this test are shown in Figure 12.2. Computation of the exact p-value is illustrated in Figure 12.3. The results from the ANOVA are typically displayed in an ANOVA table, as in Table 12.2.

FIGURE 12.2 Acceptance and rejection regions for the overall F test for one-way ANOVA

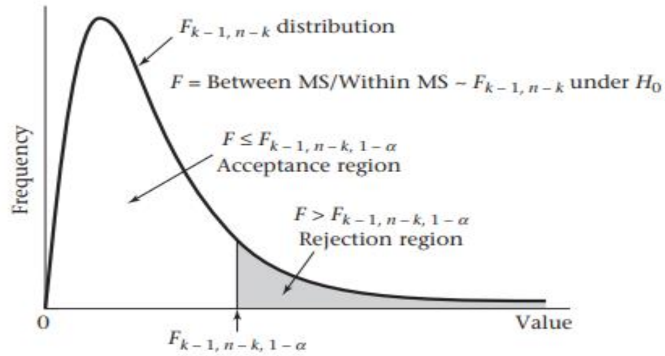


FIGURE 12.3 Computation of the exact p-value for the overall F test for one-way ANOVA

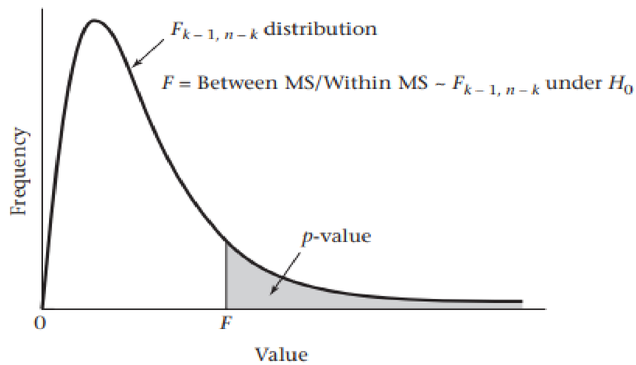


TABLE 12.2 Display of one-way ANOVA results

Source of variation	SS	df	MS	F statistic	p-value
Between	$\sum_{i=1}^{k} n_i \bar{y}_i^2 - \frac{y_{..}^2}{n} = B$	$k-1$	$\frac{B}{k-1}$	$\frac{B/(k-1)}{A/(n-k)} = F$	$Pr(F_{k-1,n-k} > F)$
Within	$\sum_{i=1}^{k} (n_i - 1)s_i^2 = A$	$n-k$	$\frac{A}{n-k}$		
Total	Between SS + Within SS				

Example (1)

A medical professional would like to know whether there is a difference in the average time it takes a patient to draw the blood sample needed to diagnose the disease in three laboratories ($k = 3$) in one of the Jordanian hospitals. The observed data, waiting times (*in minutes*), are shown in the table below:

Group (Hospital Laboratories)	Observations					
	y_{i1}	y_{i2}	y_{i3}	y_{i4}	y_{i5}	y_{i6}
Laboratory A $(i = 1)$	3	9	5	2		
Laboratory B $(i = 2)$	5	8	9	6	2	5
Laboratory C $(i = 3)$	6	3	4	5	1	

At $\alpha = 0.05$, test the claim that there is a significant difference in the mean waiting times of patients for each laboratory?

Solution

To test this claim, we proceed as follows:

Step (1): Calculate the average $(\bar{y}_{i.})$ and the variance (s_i^2) for each one of the three groups $(i = 1, 2, 3)$ as shown in the table below:

Group	Observations						Sample Size n_i	Total $y_{i.}$	Average $\bar{y}_{i.}$	Variance s_i^2
	y_{i1}	y_{i2}	y_{i3}	y_{i4}	y_{i5}	y_{i6}				
Laboratory A $(i = 1)$	3	9	5	2			4	19	4.75	9.58
Laboratory B $(i = 2)$	5	8	9	6	2	5	6	35	5.83	6.17
Laboratory C $(i = 3)$	6	3	4	5	1		5	19	3.80	3.70

Step (2): State the hypotheses as follows:

H₀: All means of the 3 groups are the same, that is, $\alpha_1 = \alpha_2 = \alpha_3 = 0$.

versus

H₁: At least two means of the 3 groups are not the same, that is, at least one $\alpha_i \neq 0$; $i = 1, 2, 3$.

Step (3): Find the critical value as follows:

Since $k = 3$ and $n = \sum_{i=1}^{k=3} n_i = n_1 + n_2 + n_3 = 4 + 6 + 5 = 15$, then:

- d_1 (df for numerator)$= k - 1 = 3 - 1 = 2$
- d_2 (df for denominator)$= n - k = 15 - 3 = 12$

Thus, the critical value is obtained from the F-table (*Table 8-Percentage points of the F distribution* ($F_{d1,\, d2,\, p}$)) in the Appendix page 882-883 as follows:

$$F_{(d_1,\ d_2,\ p\,=\,1-\alpha)} = F_{(k-1,\ n-k,\ 1-\alpha)} = F_{(2,\ 12,\ 1-0.05)}$$
$$= F_{(2,\ 12,\ 0.95)}$$
$$= 3.89$$

882 APPENDIX Tables

TABLE 8 Percentage points of the F distribution ($F_{d_1,d_2,p}$)

df for denominator, d_2	p	*df* for numerator, d_1: 1	2	3	4	5	6	7	8	12	24	∞
12	.90	3.18	2.81	2.61	2.48	2.39	2.33	2.28	2.24	2.15	2.04	1.90
	.95	4.75	3.89	3.49	3.26	3.11	3.00	2.91	2.85	2.69	2.51	2.30
	.975	6.55	5.10	4.47	4.12	3.89	3.73	3.61	3.51	3.28	3.02	2.72

Step (4): Calculate the test statistic value (*F-value*), using the following procedure:

(a) Compute the Within SS and Between SS for the blood pressure reduction data by using Equation 12.5 as follows:

(1) The sum of the observations across all groups ($y_{..}$) can be calculated as follows:

$$y_{..} = \sum_{i=1}^{k}\sum_{j=1}^{n_i} y_{ij} = \sum_{i=1}^{k} n_i * \ \bar{y}_i$$
$$y_{..} = \sum_{i=1}^{3}\sum_{j=1}^{n_i} y_{ij} = \sum_{i=1}^{3} n_i * \ \bar{y}_i$$
$$= n_1 * \ \bar{y}_1 + n_2 * \ \bar{y}_2 + n_3 * \ \bar{y}_3$$
$$= (4)(4.75) + (6)(5.83) + (5)(3.80)$$
$$= 72.98 \approx 73$$

(2) The Between Sum of Squares (Between SS) can be calculated as follows:

$$\text{Between SS} = \sum_{i=1}^{k=3} n_i\, \bar{y}_i^2 - \frac{y_{..}^2}{n}$$
$$= [(4)(4.75)^2 + (6)(5.83)^2 + (5)(3.80)^2] - \frac{(72.98)^2}{15}$$
$$= 366.3834 - 355.0720$$
$$= 11.3114$$

(3) The Within Sum of Squares (Within SS) can be calculated as follows:

$$\text{Within SS} = \sum_{i=1}^{k=3} (n_i - 1)\ s_i^2$$
$$= (n_1 - 1)\ s_1^2 + (n_2 - 1) s_2^2 + (n_3 - 1)\ s_3^2$$
$$= (3)(9.58) + (5)(6.17) + (4)(3.70)$$
$$= 74.39$$

(b) Compute the Within MS and Between MS for the blood pressure reduction data as follows:

(1) Between SS = 160.133, then:

Between MS = Between SS $/(k-1)$

= 11.3114 / 2

= 5.6557

(2) Within SS = 104.8, then:

Within MS = Within SS $/(n-k)$

= 74.39 / (15 - 3)

= 74.39 / 12

= 6.1992

(c) The test statistic value (calculated *F-value*) is obtained as follows:

F = Between MS / Within MS

= 5.6557 / 6.1992

= 0.912 $\sim F_{(k-1,\ n-k,\ 1-\alpha)} = F_{(2,\ 12,\ 0.95)}$ under H_0.

(d) The exact $p-value$ (given by the area to the right of F under H_0 $F_{(k-1,\ n-k,\ 1-\alpha)}$ distribution) can be calculated as follows:

$$\begin{aligned} p-value &= P\left(F_{(k-1,\ n-k,\ 1-\alpha)} > F\right) \\ &= P\left(F_{(2,\ 12,\ 0.95)} > 0.912\right) \\ &= 1 - 0.572 \\ &= 0.428 > \alpha = 0.05 \end{aligned}$$

(e) One-Way ANOVA Table

The results obtained in (a) – (c) are displayed in an ANOVA table (*One-Way ANOVA Table*) which is shown below:

One-Way ANOVA Table

Source of Variation	SS	df	MS	F-value	p-value
Between	11.3114	2	5.6557	0.912	0.428
Within	74.39	12	6.1992		
Total	85.7014	14			

Step (5): Make the decision. The decision is to accept (not to reject) the null hypothesis (H_0), since we get $F - value = 0.912 < F_{(2,\ 12,\ 0.95)} = 3.89$.

Step (6): Conclusion and summarizes the results. There is not enough evidence to support the claim that there is a difference among the means and conclude that the time it takes a patient to draw the blood sample needed to diagnose the disease in the three laboratories are approximately same.

12.4 Comparisons of Specific Groups in One-Way ANOVA

In the previous section (Section 12.3) a test of the hypothesis H_0: all group means are equal, versus H_1: at least two group means are different, was presented. **This test lets us detect when at least two groups have different underlying means, but it does not let us determine which of the groups have means that differ from each other**. The usual practice is to perform the overall F test just discussed. If H_0 is rejected, then specific groups are compared, as discussed in this section.

12.4.1 The t-Test for Comparison of Pairs of Groups

Suppose at this point we want to test whether groups 1 and 2 have means that are significantly different from each other.

The test statistic Z would follow an N(0, 1) distribution under H_0. Because σ^2 is generally unknown, the best estimate of it, denoted by s^2, is substituted, and the test statistic is revised accordingly. Question: How should σ^2 be estimated?

The one-way ANOVA, there are k sample variances and a similar approach is used to estimate σ^2 by computing a weighted average of k individual sample variances, where the weights are the number of degrees of freedom in each of the k samples. This formula is given as follows:

Equation 12.11

Pooled Estimate of the Variance for One-Way ANOVA

$$s^2 = \sum_{i=1}^{k}(n_i - 1)s_i^2 \Big/ \sum_{i=1}^{k}(n_i - 1) = \left[\sum_{i=1}^{k}(n_i - 1)s_i^2\right] \Big/ (n - k) = \text{Within MS}$$

Thus, the Within MS is used to estimate σ^2. The pooled estimate of the variance, that is, s^2, for the one-way ANOVA, has the following number of degrees of freedom (df):

$$(n_1 - 1) + (n_2 - 1) + \cdots + (n_k - 1) df = (n_1 + n_2 + \cdots + n_k) - k = n - k \, df$$

This test is often referred to as the least significant difference (LSD) method. The test procedure (*Least Significant Difference (LSD)*) is given as follows:

Equation 12.12

***t* Test for the Comparison of Pairs of Groups in One-Way ANOVA (LSD Procedure)**

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

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

which follows a t_{n-k} distribution under H_0.

(3) For a two-sided level α test,

if $t > t_{n-k,1-\alpha/2}$ or $t < t_{n-k,\alpha/2}$

then reject H_0

if $t_{n-k,\alpha/2} \leq t \leq t_{n-k,1-\alpha/2}$

then accept H_0

(4) The exact p-value is given by

$p = 2 \times$ the area to the left of t under a t_{n-k} distribution if $t < 0$
$= 2 \times Pr(t_{n-k} < t)$

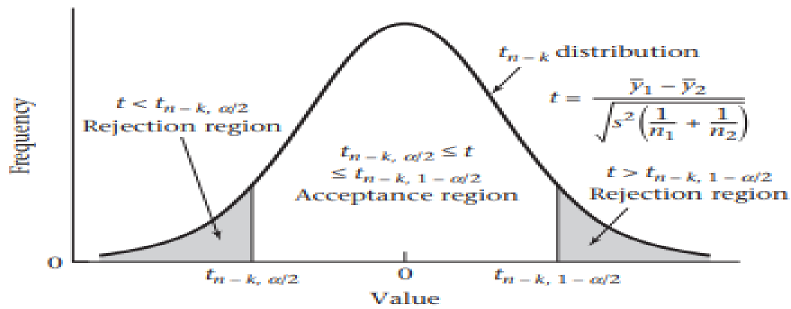
$p = 2 \times$ the area to the right of t under a t_{n-k} distribution if $t \geq 0$
$= 2 \times Pr(t_{n-k} > t)$

(5) A $100\% \times (1 - \alpha)$ CI for $\mu_1 - \mu_2$ is given by

$$\bar{y}_1 - \bar{y}_2 \pm t_{n-k,1-\alpha/2}\sqrt{s^2\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}$$

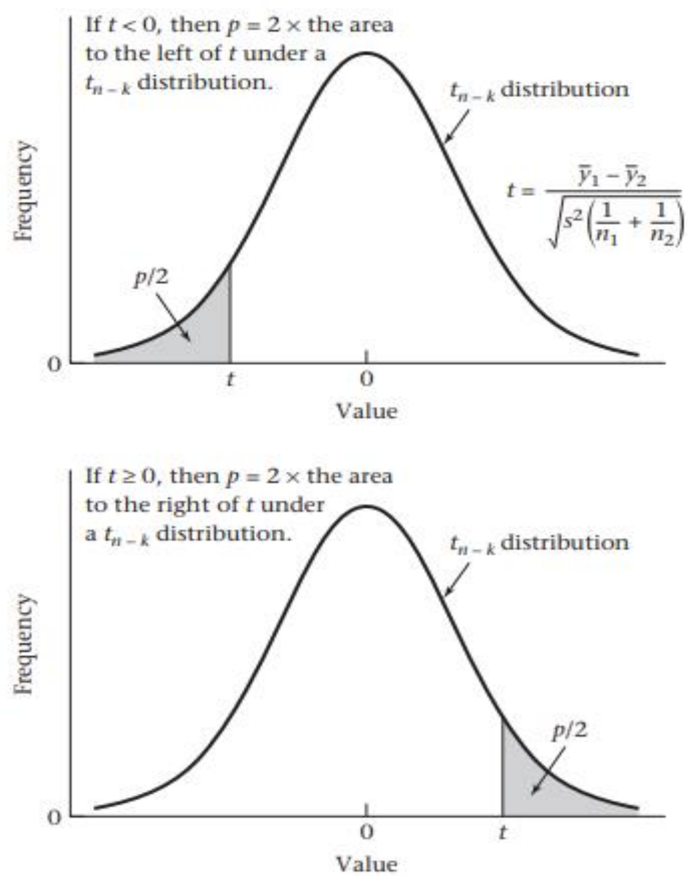
The acceptance and rejection regions for this test are given in Figure 12.4.

FIGURE 12.4 Acceptance and rejection regions for the t test for the comparison of pairs of groups in one-way ANOVA (LSD approach)



The computation of the exact p-value for the least significant difference (LSD) method is illustrated in Figure 12.5.

FIGURE 12.5 Computation of the exact p-value for the t test for the comparison of pairs of groups in one-way ANOVA (LSD approach)



Notation

The standard error (se) for an individual group mean is estimated by the formula $se = s/\sqrt{n_i}$, where $s^2 =$ Within MS.

12.4.2 Confidence Interval Method

It can also be interesting to find the $(1-\alpha)\times 100\%$ confidence intervals (CI) for the difference between two group means, say, $(\alpha_i - \alpha_j)$, for example $(\mu_1 - \mu_2)$, as follows:

A $100\% \times (1-\alpha)$ CI for $\mu_1 - \mu_2$ is given by

$$\bar{y}_1 - \bar{y}_2 \pm t_{n-k,1-\alpha/2}\sqrt{s^2\left(\frac{1}{n_1}+\frac{1}{n_2}\right)}$$

and then the following rule can be used:

(a) If ZERO (0) belongs to (∈) the $(1-\alpha)\times 100\%$ confidence intervals (CI), then we conclude that there is no difference in means for the two groups.

(b) If ZERO (0) does not belong to (∉) the $(1-\alpha)\times 100\%$ confidence intervals (CI), then we conclude that there is a difference in means for the two groups.

Example (2)

Blood Pressure: A researcher wishes to try three different techniques to lower the blood pressure of individuals diagnosed with high blood pressure. The subjects are randomly assigned to three groups $(k = 3)$ as follows:

- First group takes medication (M).
- Second group exercises (E).
- Third group diets (D).

After four weeks, the reduction in each person's blood pressure is recorded. The observed data is given as follows:

Group (*Reduction Technique*)	Observations				
	y_{i1}	y_{i2}	y_{i3}	y_{i4}	y_{i5}
Medication $(i = 1)$	10	12	9	15	13
Exercise $(i = 2)$	6	8	3	0	2
Diet $(i = 3)$	5	9	12	8	4

Answer the following:

(I) At $\alpha = 0.05$, test the claim that there is no difference among the means?

Solution

To test this claim, we proceed as follows:

Step (1): Calculate the average ($\bar{y}_{i.}$) and the variance (s_i^2) for each one of the three groups ($i = 1, 2, 3$) as shown in the table below:

Group	Observations					Totals	Averages	Variances
	y_{i1}	y_{i2}	y_{i3}	y_{i4}	y_{i5}	$y_{i.}$	$\bar{y}_{i.}$	s_i^2
Medication ($i = 1$)	10	12	9	15	13	59	11.8	5.7
Exercise ($i = 2$)	6	8	3	0	2	19	3.8	10.2
Diet ($i = 3$)	5	9	12	8	4	38	7.6	10.3

Step (2): State the hypotheses and identify the claim:

H_0: All means of blood pressure reduction observations of the 3 groups is the same, that is, $\alpha_1 = \alpha_2 = \alpha_3 = 0$ (claim).

versus

H_1: At least two means of blood pressure reduction observations of the 3 groups is not the same, that is, at least one $\alpha_i \neq 0$; $i = 1, 2, 3$.

Step (3): Find the critical value as follows:

Since $k = 3$ and $n = \sum_{i=1}^{k=3} n_i = n_1 + n_2 + n_3 = 5 + 5 + 5 = 15$, then:

- d_1 (df for numerator)$= k - 1 = 3 - 1 = 2$
- d_2 (df for denominator)$= n - k = 15 - 3 = 12$

Thus, the critical value is obtained from the F-table (*Table 8-Percentage points of the F distribution* ($F_{d1, d2, p}$)) in the Appendix page 882-883 as follows:

$$F_{(d_1,\ d_2,\ p=1-\alpha)} = F_{(k-1,\ n-k,\ 1-\alpha)} = F_{(2,\ 12,\ 1-0.05)}$$
$$= F_{(2,\ 12,\ 0.95)}$$
$$= 3.89$$

882 APPENDIX Tables

TABLE 8 Percentage points of the F distribution ($F_{d_1,d_2,p}$)

df for denominator, d_2	p	df for numerator, d_1: 1	2	3	4	5	6	7	8	12	24	∞
12	.90	3.18	2.81	2.61	2.48	2.39	2.33	2.28	2.24	2.15	2.04	1.90
	.95	4.75	3.89	3.49	3.26	3.11	3.00	2.91	2.85	2.69	2.51	2.30
	.975	6.55	5.10	4.47	4.12	3.89	3.73	3.61	3.51	3.28	3.02	2.72

Step (4): Calculate the test statistic value (*F-value*), using the following procedure:

(a) Compute the Within SS and Between SS for the blood pressure reduction data by using Equation 12.5 as follows:

(1) The sum of the observations across all groups ($y_{..}$) can be calculated as follows:

$$y_{..} = \sum_{i=1}^{k} \sum_{j=1}^{n_i} y_{ij} = \sum_{i=1}^{k} n_i * \ \bar{y}_i$$

$$\begin{aligned} y_{..} &= \sum_{i=1}^{3} \sum_{j=1}^{n_i} y_{ij} = \sum_{i=1}^{3} n_i * \ \bar{y}_i \\ &= n_1 * \ \bar{y}_1 + n_2 * \ \bar{y}_2 + n_3 * \ \bar{y}_3 \\ &= (5)(11.8) + (5)(3.8) + (5)(7.6) \\ &= 116 \end{aligned}$$

(2) The Between Sum of Squares (Between SS) can be calculated as follows:

$$\begin{aligned} \text{Between SS} &= \sum_{i=1}^{k=3} n_i \, \bar{y}_i^2 - \frac{y_{..}^2}{n} \\ &= [(5)(11.8)^2 + (5)(3.8)^2 + (5)(7.6)^2] - \frac{(116)^2}{15} \\ &= 1057.2 - 897.067 \\ &= 160.133 \end{aligned}$$

(3) The Within Sum of Squares (Within SS) can be calculated as follows:

$$\begin{aligned} \text{Within SS} &= \sum_{i=1}^{k=3} (n_i - 1) \, s_i^2 \\ &= (n_1 - 1) \, s_1^2 + (n_2 - 1) s_2^2 + (n_3 - 1) \, s_3^2 \\ &= (4)(5.7) + (4)(10.2) + (4)(10.3) = 104.8 \end{aligned}$$

(b) Compute the Within MS and Between MS for the blood pressure reduction data as follows:

(1) Between SS = 160.133, then:

Between MS = Between SS /$(k - 1)$

= 160.133 / 2

= 80.0665

(2) Within SS = 104.8, then:

Within MS = Within SS /$(n - k)$

= 104.8 / (15 - 3)

= 104.8 / 12

= 8.7333

(c) The test statistic value (calculated *F-value*) is obtained as follows:

F = Between MS / Within MS

= 80.0665 / 8.7333

= 9.17 $\sim F_{(k-1,\ n-k,\ 1-\alpha)} = F_{(2,\ 12,\ 0.95)}$ under H_0.

(d) The exact $p-value$ (given by the area to the right of F under an $F_{(k-1,\ n-k,\ 1-\alpha)}$ distribution) can be calculated as follows:

$$p-value = P\left(F_{(k-1,\ n-k,\ 1-\alpha)} > F\right)$$
$$= P\left(F_{(2,\ 12,\ 0.95)} > 9.17\right)$$
$$= 0.004 < \alpha = 0.05$$

(d) One-Way ANOVA Table

The results obtained in (a) – (c) are displayed in an ANOVA table (*One-Way ANOVA Table*) which is shown below:

One-Way ANOVA Table

<table>
<tr><th>Source of Variation</th><th>SS</th><th>df</th><th>MS</th><th>F-value</th><th>p-value</th></tr>
<tr><td>Between</td><td>160.133</td><td>2</td><td>80.0665</td><td rowspan="3">9.17</td><td rowspan="3">0.004</td></tr>
<tr><td>Within</td><td>104.8</td><td>12</td><td>8.7333</td></tr>
<tr><td>Total</td><td>264.933</td><td>14</td><td></td></tr>
</table>

Step (5): Make the decision. The decision is to reject the null hypothesis (H_0), since we get $F-value = 9.17 > F_{(2,\ 12,\ 0.95)} = 3.89$.

Step (6): Conclusion and summarizes the results. There is enough evidence to reject the claim and conclude that at least one mean is different from the others.

(II) At $\alpha = 0.05$, use the least significant difference (LSD) method to determine specific differences between blood pressure reduction techniques?

Solution

Science the decision in part (I) indicates that a difference exists between the means of the blood pressure reduction techniques (*because we reject H_0*), then we will perform the least significant difference (LSD) method to isolate the specific difference.

Step (1): Critical Value

The critical value will be obtained from Table 5 in the Appendix based on degrees of freedom $df = n - k = 15 - 3 = 12$, as follows:

$$t_{(n-k\,,\ 1-\alpha/2)} = t_{(15-3\,,\ 1-0.05/2)} = t_{(12\,,\ 0.975)} = 2.179$$

Step (2): $s^2 = \text{Within MS} = 8.7333$

Step (3): The value of the test statistic (t) for the all pairs of compared groups is calculated as follows:

(a) Groups Compared - Medication (M) and Exercise (E):

Hypothesis: H_0: $\alpha_M = \alpha_E$ $versus$ H_1: $\alpha_M \neq \alpha_E$

$$t = \frac{\bar{y}_M - \bar{y}_E}{\sqrt{s^2\left(\frac{1}{n_M} + \frac{1}{n_E}\right)}} = \frac{11.8 - 3.8}{\sqrt{8.7333\left(\frac{1}{5} + \frac{1}{5}\right)}} = \frac{8}{1.869} = 4.280$$

(b) Groups Compared - Medication (M) and Diet (D):

Hypothesis: H_0: $\alpha_M = \alpha_D$ $versus$ H_1: $\alpha_M \neq \alpha_D$

$$t = \frac{\bar{y}_M - \bar{y}_D}{\sqrt{s^2\left(\frac{1}{n_M} + \frac{1}{n_E}\right)}} = \frac{11.8 - 7.6}{\sqrt{8.7333\left(\frac{1}{5} + \frac{1}{5}\right)}} = \frac{4.2}{1.869} = 2.247$$

(c) Groups Compared - Exercise (E) and Diet (D):

Hypothesis: H_0: $\alpha_E = \alpha_D$ $versus$ H_1: $\alpha_E \neq \alpha_D$

$$t = \frac{\bar{y}_E - \bar{y}_D}{\sqrt{s^2\left(\frac{1}{n_M} + \frac{1}{n_E}\right)}} = \frac{3.8 - 7.6}{\sqrt{8.7333\left(\frac{1}{5} + \frac{1}{5}\right)}} = \frac{-3.8}{1.869} = -2.033$$

Therefore, the results of the comparisons using the LSD method are presented in the following table:

Groups Compared	**Test Statistic Value**	**Critical Value**	**Decision**
Medication (M) , Exercise (E)	4.280	2.179	Reject H_0
Medication (M) , Diet (D)	2.247	2.179	Reject H_0
Exercise (E) , Diet (D)	– 2.033	2.179	Accept (Do Not Reject) H_0

Step (4): Conclusion

There are no significant differences (t = -2.033 < $t_{(12\,,\;0.975)} = 2.179$) between the Exercise (E) and Diet (D) means. Both techniques, Exercise and Diet, have approximately the same effect on lowering the blood pressure of individuals diagnosed with high blood pressure.

(III) Find a 95% confidence intervals for the difference between the mean blood pressure reduction for all techniques $\mu_M - \mu_E$, $\mu_M - \mu_D$ and $\mu_E - \mu_D$?

Solution

The $(1-\alpha)\times 100\%$ confidence intervals (CI) for the difference between two group means, say,$(\alpha_i - \alpha_j)$, for example $(\mu_1 - \mu_2)$, can be obtained as follows:

A $100\% \times (1-\alpha)$ CI for $\mu_1 - \mu_2$ is given by

$$\bar{y}_1 - \bar{y}_2 \pm t_{n-k,1-\alpha/2}\sqrt{s^2\left(\frac{1}{n_1}+\frac{1}{n_2}\right)}$$

Step (1): $s^2 = \text{Within MS} = 8.7333$

Step (2): $t_{(12\,,\;0.975)} = 2.179$

Step (3): The 95% confidence interval for $\mu_M - \mu_E$ is given by:

$$\text{CI} = (\bar{y}_M - \bar{y}_E) \pm t_{(n-k\,,\;1-\alpha/2)}\sqrt{s^2\left(\frac{1}{n_M}+\frac{1}{n_E}\right)}$$

$$= (11.8 - 3.8) \pm 2.179\sqrt{8.7333 * \left(\frac{1}{5}+\frac{1}{5}\right)} \;= 8 \pm 4.073 = (3.93\,, 12.07)$$

Conclusion

We conclude that $0 \notin \text{CI} = (3.93\,, 12.07)$ which implies that there is a difference in means for the two techniques, Medication (M) and Exercise (E), on lowering the blood pressure of individuals diagnosed with high blood pressure.

Step (4): The 95% confidence interval for $\mu_M - \mu_D$ is given by:

$$\text{CI} = (\bar{y}_M - \bar{y}_D) \pm t_{(n-k\,,\ 1-\alpha/2)} \sqrt{s^2 \left(\frac{1}{n_M} + \frac{1}{n_D}\right)}$$

$$= (11.8 - 7.6) \pm 2.179 \sqrt{8.7333 * \left(\tfrac{1}{5} + \tfrac{1}{5}\right)} = 4.2 \pm 4.073 = (0.13\,, 8.27)$$

Conclusion

We conclude that 0 $\notin \text{CI} = (0.13\,, 8.27)$ which implies that there is a difference in means for the two techniques, Medication (M) and Diet (D), on lowering the blood pressure of individuals diagnosed with high blood pressure.

Step (5): The 95% confidence interval for $\mu_E - \mu_D$ is given by:

$$\text{CI} = (\bar{y}_E - \bar{y}_D) \pm t_{(n-k\,,\ 1-\alpha/2)} \sqrt{s^2 \left(\frac{1}{n_E} + \frac{1}{n_D}\right)}$$

$$= (3.8 - 7.6) \pm 2.179 \sqrt{8.7333 * \left(\tfrac{1}{5} + \tfrac{1}{5}\right)} = -3.8 \pm 4.073 = (-7.87\,, 0.27)$$

Conclusion

We conclude that 0 $\in \text{CI} = (-7.87\,, 0.27)$ which implies that there is no difference in means for the two techniques, Exercise (E) and Diet (D), on lowering the blood pressure of individuals diagnosed with high blood pressure.

Therefore, the results of the 95% confidence intervals for the diferences between the mean blood pressure reduction for all reduction techniques $\mu_M - \mu_E$, $\mu_M - \mu_D$ and $\mu_E - \mu_D$ are presented in the following table:

Mean Differences	**95% Confidence Interval**		**Includes Zero**	**Decision**
	Lower Limit	Upper Limit		
$\mu_M - \mu_E$	3.93	12.07	No	Reject H_0
$\mu_M - \mu_D$	0.13	8.27	No	Reject H_0
$\mu_E - \mu_D$	-7.87	0.27	Yes	Accept (Do Not Reject) H_0

Exercises

Exercise (1)

A pharmaceutical company conducts an experiment to test the effect of a new cholesterol medication. The company selects 15 subjects randomly from a larger population ($n = 15$). Each subject is randomly assigned to one of three treatment groups ($k = 3$). Within each treatment group, subjects receive a different dose of the new medication. In Group 1, subjects receive 0 mg/day; in Group 2, subjects receive 50 mg/day; and in Group 3, subjects receive 100 mg/day. After 30 days, doctors measure the cholesterol level of each subject. The results for all 15 subjects appear in the table below:

Groups	Observations				
	y_{i1}	y_{i2}	y_{i3}	y_{i4}	y_{i5}
Group 1, 0 mg $(i = 1)$	210	240	270	270	300
Group 2, 50 mg $(i = 2)$	210	240	240	270	270
Group 3, 100 mg $(i = 3)$	180	210	210	210	240

At $\alpha = 0.01$, test if there is a significant effect of the dosage level on the cholesterol level?

Answer

To make this test, we proceed as follows:

Step (1): Basic Calculations as follows:

Calculate the average ($\bar{y}_{i.}$) and the variance (s_i^2) for each one of the three groups ($i = 1, 2, 3$) as shown in the table given below:

Group	Observations					Sample Size n_i	Total $y_{i.}$	Average $\bar{y}_{i.}$	Variance s_i^2
	y_{i1}	y_{i2}	y_{i3}	y_{i4}	y_{i5}				
Group 1, 0 mg $(i = 1)$	210	240	270	270	300	5	1290	258	1170
Group 2, 50, mg $(i = 2)$	210	240	240	270	270	5	1230	246	630
Group 3, 100 mg $(i = 3)$	180	210	210	210	240	5	1050	210	450

Step (2): State the hypotheses as follows:

H$_0$: All means of the 3 groups are the same, that is, $\alpha_1 = \alpha_2 = \alpha_3 = 0$.

$versus$

H$_1$: At least two means of the 3 groups are not the same, that is, at least one $\alpha_i \neq 0$; $i = 1, 2, 3$.

Step (3): Find the critical value as follows:

Since $k = 3$ and $n = \sum_{i=1}^{k=3} n_i = n_1 + n_2 + n_3 = 5 + 5 + 5 = 15$, then:

- d_1 (df for numerator)$= k - 1 = 3 - 1 = 2$
- d_2 (df for denominator)$= n - k = 15 - 3 = 12$

Thus, the critical value is obtained from the F-table (*Table 8-Percentage points of the F distribution* ($F_{d1, d2, p}$)) in the Appendix page 882-883 as follows:

$$F_{(d_1,\ d_2,\ p=1-\alpha)} = F_{(k-1,\ n-k,\ 1-\alpha)} = F_{(2,\ 12,\ 1-0.01)}$$
$$= F_{(2,\ 12,\ 0.99)}$$
$$= 6.93$$

Step (4): Calculate the test statistic value (*F-value*), using the following procedure:

(a) Compute the Within SS and Between SS for the blood pressure reduction data by using Equation 12.5 as follows:

(1) The sum of the observations across all groups ($y_{..}$) can be calculated as follows:

$$y_{..} = \sum_{i=1}^{k}\sum_{j=1}^{n_i} y_{ij} = \sum_{i=1}^{k} n_i * \bar{y}_i$$
$$y_{..} = \sum_{i=1}^{3}\sum_{j=1}^{n_i} y_{ij} = \sum_{i=1}^{3} n_i * \bar{y}_i$$
$$= n_1 * \bar{y}_1 + n_2 * \bar{y}_2 + n_3 * \bar{y}_3$$
$$= (5)(258) + (5)(246) + (5)(210)$$
$$= 3570$$

(2) The Between Sum of Squares (Between SS) can be calculated as follows:

$$\text{Between SS} = \sum_{i=1}^{k=3} n_i\, \bar{y}_i^2 - \frac{y_{..}^2}{n}$$
$$= [(5)(258)^2 + (5)(246)^2 + (5)(210)^2] - \frac{(3570)^2}{15}$$
$$= 855900 - 849660$$
$$= 6240$$

(3) The Within Sum of Squares (Within SS) can be calculated as follows:

$$
\begin{aligned}
\text{Within SS} &= \sum_{i=1}^{k=3}(n_i - 1)\, s_i^2 \\
&= (n_1 - 1)\, s_1^2 + (n_2 - 1) s_2^2 + (n_3 - 1)\, s_3^2 \\
&= (4)(1170) + (4)(630) + (4)(450) \\
&= 9000
\end{aligned}
$$

(b) Compute the Within MS and Between MS for the blood pressure reduction data as follows:

(1) Between SS = 6240, then:

Between MS = Between SS $/(k-1)$

= 6240 / 2

= 3120

(2) Within SS = 9000, then:

Within MS = Within SS $/(n-k)$

= 9000 / (15 - 3)

= 9000 / 12

= 750

(c) The test statistic value (calculated *F-value*) is obtained as follows:

F = Between MS / Within MS

= 3120 / 750

= 4.16 $\sim F_{(k-1,\ n-k,\ 1-\alpha)} = F_{(2,\ 12,\ 0.99)}$ under H_0.

(d) The exact $p - value$ (given by the area to the right of F under H_0 $F_{(k-1,\ n-k,\ 1-\alpha)}$ distribution) can be calculated as follows:

$$
\begin{aligned}
p - value &= P\left(F_{(k-1,\ n-k,\ 1-\alpha)} > F\right) \\
&= P\left(F_{(2,\ 12,\ 0.99)} > 4.16\right) \\
&= 1 - P\left(F_{(2,\ 12,\ 0.99)} \leq 4.16\right) \\
&= 1 - 0.958 \\
&= 0.042 > \alpha = 0.01
\end{aligned}
$$

(e) One-Way ANOVA Table

The results obtained in (a) – (c) are displayed in an ANOVA table (*One-Way ANOVA Table*) which is shown below:

One-Way ANOVA Table

Source of Variation	SS	df	MS	F-value	p-value
Between	6240	2	3120	4.16	0.042
Within	9000	12	750		
Total	15240	14			

Step (5): Make the decision. The decision is to accept (not to reject) the null hypothesis (H_0), since we get $F - value = 4.16 < F_{(2,\ 12,\ 0.99)} = 6.93$.

Step (6): Conclusion and summarizes the results. There is not enough evidence to support the claim that there is a difference among the means and conclude that there is no effect of the dosage level on the cholesterol level. The effect of the three dosage levels on the cholesterol level is approximately the same.

Exercise (2)

A researcher wishes to see whether there is any difference in the weight gains of athletes following one of three special diets. Athletes are randomly assigned to three groups and placed on the diet for six weeks. The weight gains (*in pounds*) are shown in the table given below:

Diet		
A	B	C
16	18	26
15	22	31
13	20	24
21	16	30
15	24	24

Answer the following:

(I) At $\alpha = 0.01$, can the researcher conclude that there is a difference in the diets?

Answer

One-Way ANOVA Table

Source of Variation	SS	df	MS	F-value	p-value
Between	310	2	155	15.5	0.000
Within	120	12	10		
Total	430	14			

Decision and Conclusion: The decision is to reject the null hypothesis (H_0), since we get $F - value = 15.5 > F_{(2,\ 12,\ 0.99)} = 6.93$. There is evidence, at the 1% significance level, that the true mean of weight gains of athletes of the three special diet groups is different. Therefore, the researcher can conclude that there is a difference in the three special diets programs.

(II) At $\alpha = 0.01$, use the least significant difference (LSD) method to determine specific differences between special diets programs?

Answer

Science the decision in part (I) indicates that a difference exists between the mean of weight gains of athletes of the three diet groups (*because we reject H_0*), then we will perform the least significant difference (LSD) method to isolate the specific difference as follows:

(1) Critical value: $t_{(n-k\,,\ \alpha/2)} = t_{(15-3\,,\ 0.01/2)} = t_{(12\,,\ 0.005)} = -3.055$

(2) $s^2 =$ Within MS $= 10$

(3) The value of the test statistic (t) for the all pairs of compared groups is calculated as follows:

(a) Groups Compared - Diet (A) and Diet (B):

Hypothesis: H_0: $\alpha_A = \alpha_B$ *versus* H_1: $\alpha_A \neq \alpha_B$

$$t = \frac{\bar{y}_A - \bar{y}_B}{\sqrt{s^2\left(\frac{1}{n_A} + \frac{1}{n_B}\right)}} = \frac{16 - 20}{\sqrt{10\left(\frac{1}{5} + \frac{1}{5}\right)}} = \frac{-4}{2} = -2$$

(b) Groups Compared - Diet (A) and Diet (C):

Hypothesis: H_0: $\alpha_A = \alpha_C$ *versus* H_1: $\alpha_A \neq \alpha_C$

$$t = \frac{\bar{y}_A - \bar{y}_C}{\sqrt{s^2\left(\frac{1}{n_A} + \frac{1}{n_C}\right)}} = \frac{16 - 27}{\sqrt{10\left(\frac{1}{5} + \frac{1}{5}\right)}} = \frac{-11}{2} = -5.5$$

(c) Groups Compared - Diet (B) and Diet (C):

Hypothesis: H_0: $\alpha_B = \alpha_C$ *versus* H_1: $\alpha_B \neq \alpha_C$

$$t = \frac{\bar{y}_B - \bar{y}_C}{\sqrt{s^2\left(\frac{1}{n_B} + \frac{1}{n_C}\right)}} = \frac{20 - 27}{\sqrt{10\left(\frac{1}{5} + \frac{1}{5}\right)}} = \frac{-7}{2} = -3.5$$

Therefore, the results of the comparisons using the LSD method are presented in the following table:

Groups Compared	Test Statistic Value	Critical Value	Decision
Diet (A), Diet (B)	– 2	-3.055	Accept (Do Not Reject) H_0
Diet (A), Diet (C)	– 5.5	-3.055	Reject H_0
Diet (B), Diet (C)	– 3.5	-3.055	Reject H_0

(4) Conclusion: There are no significant differences (t = -2 > $t_{(12,\ 0.005)} = -3.055$) between the Diet (A) and Diet (B) means. Both special diet programs, Diet (A) and Diet (B), have approximately the same effect on the weight gains of athletes.

(III) Find a 99% confidence intervals for the difference between the mean of weight gains of athletes of the three special diet programs, that is, $\mu_A - \mu_B$, $\mu_A - \mu_C$ and $\mu_B - \mu_C$?

Solution

The $(1 - \alpha) \times 100\%$ confidence intervals (CI) for the difference between two group means, say, $(\alpha_i - \alpha_j)$, for example $(\mu_1 - \mu_2)$, can be obtained as follows:

A 100% × (1 − α) CI for $\mu_1 - \mu_2$ is given by

$$\bar{y}_1 - \bar{y}_2 \pm t_{n-k,1-\alpha/2}\sqrt{s^2\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}$$

Step (1): $s^2 =$ Within MS $= 10$

Step (2): $t_{(12\,,\ 0.975)} = 3.055$

Step (3): The 99% confidence interval for $\mu_A - \mu_B$ is given by:

$$\text{CI} = (\bar{y}_A - \bar{y}_B) \pm t_{(n-k\,,\ 1-\alpha/2)}\sqrt{s^2\left(\frac{1}{n_A} + \frac{1}{n_B}\right)}$$

$$= (16 - 20) \pm 3.055\sqrt{10 * \left(\tfrac{1}{5} + \tfrac{1}{5}\right)} = -4 \pm 6.11 = (-10.11\,, 2.11)$$

Conclusion

We conclude that $0 \in \text{CI} = (-10.11\,, 2.11)$ which implies that there is no difference in means for the two diets, Diet (A) and Diet (B), on weight gains of athletes.

Step (4): The 99% confidence interval for $\mu_A - \mu_C$ is given by:

$$\text{CI} = (\bar{y}_A - \bar{y}_C) \pm t_{(n-k\,,\ 1-\alpha/2)}\sqrt{s^2\left(\frac{1}{n_A} + \frac{1}{n_C}\right)}$$

$$= (16 - 27) \pm 3.055\sqrt{10 * \left(\tfrac{1}{5} + \tfrac{1}{5}\right)} = -11 \pm 6.11 = (-17.11\,, -4.89)$$

Conclusion

We conclude that $0 \notin \text{CI} = (-17.11\,, -4.89)$ which implies that there is a difference in means for the two diets, Diet (A) and Diet (C), on weight gains of athletes.

Step (5): The 99% confidence interval for $\mu_B - \mu_C$ is given by:

$$\text{CI} = (\bar{y}_B - \bar{y}_C) \pm t_{(n-k\,,\ 1-\alpha/2)}\sqrt{s^2\left(\frac{1}{n_B} + \frac{1}{n_C}\right)}$$

$$= (20 - 27) \pm 3.055\sqrt{10 * \left(\tfrac{1}{5} + \tfrac{1}{5}\right)} = -7 \pm 6.11 = (-13.11\,, -0.89)$$

Conclusion

We conclude that $0 \notin \text{CI} = (-13.11\,, -0.89)$ which implies that there is a difference in means for the two diets, Diet (B) and Diet (C), on weight gains of athletes.

Therefore, the results of the 99% confidence intervals for the diferences between the mean blood pressure reduction for all reduction techniques $\mu_M - \mu_E$, $\mu_M - \mu_D$ and $\mu_E - \mu_D$ are presented in the following table:

Mean Differences	99% Confidence Interval		Includes Zero	Decision
	Lower Limit	Upper Limit		
$\mu_A - \mu_B$	-10.11	2.11	Yes	Accept (Do Not Reject) H_0
$\mu_A - \mu_C$	-17.11	-4.89	No	Reject H_0
$\mu_B - \mu_C$	-13.11	-0.89	No	Reject H_0
