

One-Way Analysis of Variance

Nathaniel E. Helwig

University of Minnesota

1 Model Form

The one-way ANalysis Of VAriance (ANOVA) model extends the independent samples t test problem to the situation with $a > 2$ groups. Suppose that we have an independent sample of observations where $y_{ij} \stackrel{\text{ind}}{\sim} N(\mu_j, \sigma^2)$ for $i = 1, \dots, n_j$ and $j = 1, \dots, a$. In other words, the one-way ANOVA model assumes that

$$y_{ij} = \mu_j + \epsilon_{ij}$$

where μ_j is the mean for the j -th group, and the error terms are iid normal variables with mean zero and homogeneous variance, i.e., $\epsilon_{ij} \stackrel{\text{iid}}{\sim} N(0, \sigma^2)$. The above form of the model is known as the “cell means” model form, given that it explicitly writes each group’s mean using the parameter μ_j . We could also write the model in the “treatment effects” model form, which decomposes the group’s means such as

$$y_{ij} = \mu + \tau_j + \epsilon_{ij}$$

where μ is the overall mean that is common to all observations, and τ_j is the j -th group’s treatment effect, which satisfies $\sum_{j=1}^a \tau_j = 0$. To connect the two different forms of the model, note that $\mu_j = \mu + \tau_j$, which reveals that $\tau_j = \mu_j - \mu$ is the j -th group’s deviation from the overall mean. The treatment effect version of the model tends to be more useful for comparing differences between the group’s, given that it explicitly parameterizes the similarity (i.e., μ parameter) and differences (i.e., τ_j parameters) between the groups.

2 Parameter Estimation

To estimate the μ_j parameters, consider minimizing the least squares loss function

$$\sum_{j=1}^a \sum_{i=1}^{n_j} (y_{ij} - \mu_j)^2 = \sum_{j=1}^a L_j(\mu_j | \mathbf{y}_j)$$

where $L_j(\cdot)$ is the least squares loss function for the j -th group, and $\mathbf{y}_j = (y_{1j}, \dots, y_{n_j j})^\top$ is the vector of observed data for the j -th group. The least squares loss function for each group can be written as

$$L_j(\mu_j | \mathbf{y}_j) = \sum_{i=1}^{n_j} (y_{ij} - \mu_j)^2 = \sum_{i=1}^{n_j} y_{ij}^2 - 2\mu_j n_j \bar{y}_j + n_j \mu_j^2$$

where $\bar{y}_j = \frac{1}{n_j} \sum_{i=1}^{n_j} y_{ij}$ is the j -th group's sample mean. Note that $L_j(\cdot)$ is the same least squares loss function that we encountered in our Parameter Estimation chapter. Taking the derivative with respect to μ_j , setting the derivative to zero, and solving for μ_j reveals that

$$\hat{\mu}_j = \bar{y}_j$$

is the least squares estimate of the j -th group's mean (which is not too surprising).

Using the treatment effect parameterization, the overall mean is estimated as

$$\hat{\mu} = \frac{1}{a} \sum_{j=1}^a \hat{\mu}_j = \frac{1}{a} \sum_{j=1}^a \bar{y}_j$$

which is simply the average of the group means. The treatment effects have the form

$$\hat{\tau}_j = \hat{\mu}_j - \hat{\mu}$$

and note that the treatment effect estimates solve the “sum of zero” constraint

$$\sum_{j=1}^a \hat{\tau}_j = \sum_{j=1}^a \hat{\mu}_j - a\hat{\mu} = 0$$

given that the overall mean is simply the average of the group means.

3 Sums of Squares

The model is referred to as an “analysis of variance” model because the model makes it possible to analyze how the variance in the data can be decomposed into the within and between group variances. The *sum of squares total* is defined as

$$\text{SST} = \sum_{j=1}^a \sum_{i=1}^{n_j} (y_{ij} - \bar{y})^2$$

where $\bar{y} = \frac{1}{n} \sum_{j=1}^a \sum_{i=1}^{n_j} y_{ij}$ is the overall mean and $n = \sum_{j=1}^a n_j$ is the total sample size. Note that if we multiplied the SST quantity by $\frac{1}{n-1}$, this would be the sample variance of the data ignoring the group membership. The SST quantity has $n - 1$ degrees of freedom, given that there are n independent values and one constraint: $\sum_{j=1}^a \sum_{i=1}^{n_j} (y_{ij} - \bar{y}) = 0$.

The *sum of square between* groups is defined as

$$\text{SSB} = \sum_{j=1}^a \sum_{i=1}^{n_j} (\bar{y}_j - \bar{y})^2 = \sum_{j=1}^a n_j (\bar{y}_j - \bar{y})^2$$

which quantifies how the group means \bar{y}_j vary around the overall mean \bar{y} . Note that if $n_j = n_k$ for all j, k , then the SSB quantity is proportional to the sample variance of the group means. The SSB quantity has $a - 1$ degrees of freedom, given that there are a independent group means and one constraint: $\sum_{j=1}^a n_j (\bar{y}_j - \bar{y}) = 0$.

Finally, the *sum of squares within* groups is defined as

$$\text{SSW} = \sum_{j=1}^a \sum_{i=1}^{n_j} (y_{ij} - \bar{y}_j)^2$$

and note that if we multiplied the SSW quantity by $\frac{1}{n-a}$ this would be the pooled estimate of the sample variance of the data conditioned on the group membership. The SSW quantity has $n - a$ degrees of freedom given that there are n independent values and a constraints: $\sum_{i=1}^{n_j} (y_{ij} - \bar{y}_j) = 0$ for $j = 1, \dots, a$.

The degrees of freedom of the three sums of square are additive, i.e.,

$$df_T = df_B + df_W$$

where $df_T = n - 1$ is the SST degrees of freedom, $df_B = a - 1$ is the SSB degrees of freedom, and $df_W = n - a$ is the SSW degrees of freedom. Although it may not be obvious from their definitions, the sums of squares can be decomposed such as

$$SST = SSB + SSW$$

which implies that the total variation (SST) is the summation of the between group variation (SSB) and the within group variation (SSW). To prove the relationship, we can write

$$\begin{aligned} SST &= \sum_{j=1}^a \sum_{i=1}^{n_j} (y_{ij} - \bar{y}_j + \bar{y}_j - \bar{y})^2 \\ &= \sum_{j=1}^a \sum_{i=1}^{n_j} (\bar{y}_j - \bar{y})^2 + \sum_{j=1}^a \sum_{i=1}^{n_j} (y_{ij} - \bar{y}_j)^2 + 2 \sum_{j=1}^a \sum_{i=1}^{n_j} (y_{ij} - \bar{y}_j) (\bar{y}_j - \bar{y}) \end{aligned}$$

where the first term is the SSB, the second term is the SSW, and the third term is zero:

$$\begin{aligned} \sum_{j=1}^a \sum_{i=1}^{n_j} (y_{ij} - \bar{y}_j) (\bar{y}_j - \bar{y}) &= \sum_{j=1}^a \sum_{i=1}^{n_j} (y_{ij} \bar{y}_j - y_{ij} \bar{y} - \bar{y}_j^2 + \bar{y}_j \bar{y}) \\ &= \sum_{j=1}^a n_j \bar{y}_j^2 - n \bar{y}^2 - \sum_{j=1}^a n_j \bar{y}_j^2 + n \bar{y}^2 = 0 \end{aligned}$$

The *coefficient of multiple determination* (also known as R-squared) is defined as

$$R^2 = \frac{SSB}{SST} = 1 - \frac{SSE}{SST}$$

which is the proportion of the total variation in the sample of data that can be accounted for by the group membership. The adjusted coefficient of multiple determination (or adjusted R-squared) value is defined as

$$R_{adj}^2 = 1 - \frac{SSW/df_W}{SST/df_T} = 1 - \frac{\hat{\sigma}^2}{s_y^2}$$

where $\hat{\sigma}^2$ is the mean squared error (defined in the next section) and s_y^2 is the sample variance of Y ignoring the group membership.

4 Omnibus F Test

Suppose that we want to test whether there are any difference between the groups' means. This corresponds to testing the null hypothesis $H_0 : \mu_j = \mu \forall j$ versus the alternative hypothesis $H_1 : (\exists j)(\mu_j \neq \mu)$. Writing this in terms of the treatment effects, the hypotheses could be rewritten as $H_0 : \tau_j = 0 \forall j$ versus $H_1 : (\exists j)(\tau_j \neq 0)$. To test the null hypothesis, we can use the F test statistic

$$F = \frac{\text{SSB}/df_B}{\text{SSW}/df_W}$$

where the numerator is known as the *mean square between* (or mean square treatment) and the denominator is referred to as the *mean square within* (or mean square error).

The mean square within is an unbiased estimate of the error variance, i.e., the quantity

$$\hat{\sigma}^2 = \text{SSW}/df_W$$

has expectation $E(\hat{\sigma}^2) = \sigma^2$, where σ^2 is the variance of the ϵ terms in the ANOVA model. The unbiasedness of $\hat{\sigma}^2$ can be proved using a similar approach that was used in the Parameter Estimation notes to show that the sample variance (dividing by $n-1$) is unbiased. Note that the sample variance $s^2 = \frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})^2$ has the same form as the mean square within, in the sense that it is a sum of squares divided by its degrees of freedom.

Assuming that H_0 is true, $F \sim F(df_B, df_W)$ where the notation $F(df_B, df_W)$ denotes an F distribution with degrees of freedom parameters (df_B, df_W) . Note that as the observed F test statistic gets larger, this provides more evidence against the null hypothesis—because large F values indicate that the variation between groups is large compared to the variation within groups. Consequently, to test the null hypothesis with a significance level of α , we only need to worry about the upper tail of the F_{df_B, df_W} distribution. Specifically, if the observed test statistic is larger than $F_{1-\alpha}(df_B, df_W)$, then we would reject H_0 . Note that $F_{1-\alpha}(df_B, df_W)$ denotes the $1 - \alpha$ quantile of the $F(df_B, df_W)$ distribution.

This is referred to as an *omnibus* test because this approach simultaneously tests the significance of all of the treatment effects. Note that if we reject the null hypothesis, this does not inform us which of the group means significantly differ from one another. As a result, it is necessary to use a different approach if your goal is to determine which means are significantly different from one another.

Example 1. Suppose a researcher wants to study how the visual/auditory synchrony of information affects one's ability to recall the information. The researcher took a random sample of $n = 30$ college students, and asked each student to watch a video of a person reciting a list of 50 words. The students were instructed to try to remember and recall as many of the words as possible. The students were randomly assigned to one of three treatment conditions with $n_j = 10$ students in each condition. The three conditions are (i) fast: sound precedes lip movements in video, (ii) normal: sound synced with lip movements in video, and (iii) slow: lip movements in video precede sound. In this case, we want to test the null hypothesis $H_0 : \mu_1 = \mu_2 = \mu_3$ versus the alternative hypothesis $H_1 : \text{not all } \mu_j \text{ are equal}$, where μ_j is the mean number of recalled words for the j -th group.

```
> words <- c(23, 27, 23, 22, 28, 24, 18, 33, 21, 15,
+           19, 25, 29, 25, 19, 30, 29, 24, 23, 36,
+           22, 16, 30, 17, 19, 26, 20, 17, 21, 23)
> cond <- factor(rep(c("fast", "normal", "slow"), 10))
> words
[1] 23 27 23 22 28 24 18 33 21 15 19 25 29 25 19 30 29 24 23 36 22 16
[23] 30 17 19 26 20 17 21 23
> cond
[1] fast    normal slow    fast    normal slow    fast    normal slow
[10] fast    normal slow    fast    normal slow    fast    normal slow
[19] fast    normal slow    fast    normal slow    fast    normal slow
[28] fast    normal slow
Levels: fast normal slow

> data <- matrix(words, ncol = 3, byrow = TRUE)
> colnames(data) <- c("fast", "normal", "slow")
> data
      fast normal slow
[1,]   23     27   23
[2,]   22     28   24
[3,]   18     33   21
[4,]   15     19   25
```

```
[5,] 29 25 19
[6,] 30 29 24
[7,] 23 36 22
[8,] 16 30 17
[9,] 19 26 20
[10,] 17 21 23
```

```
> # fit anova model
> amod <- aov(words ~ cond)
> amod
Call:
aov(formula = words ~ cond)
```

Terms:

	cond	Residuals
Sum of Squares	233.8667	535.6000
Deg. of Freedom	2	27

Residual standard error: 4.453879
 Estimated effects may be unbalanced

```
> # omnibus F test
> anova(amod)
Analysis of Variance Table
```

Response: words

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
cond	2	233.87	116.933	5.8947	0.007513 **
Residuals	27	535.60	19.837		

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

The one-way ANOVA has an observed test statistic of $F = 5.8947$, which follows an $F(2, 27)$ distribution under H_0 . The p-value is $p = P(F_{2,27} > 5.8947) = 0.007513$, which means that we can reject H_0 using any standard significance level. However, rejecting the null hypothesis $H_0 : \mu_1 = \mu_2 = \mu_3$ doesn't tell us which conditions significantly differ from one another. In other words, the omnibus F test doesn't help us understand how the different synchrony conditions affect the number of words recalled.

5 Multiple Comparisons

5.1 Overview

To understand the nature of the mean differences, we need to perform followup tests (multiple comparisons) to determine where the mean differences are occurring in the data. In most cases, we are interested in follow-up tests about linear combinations of the factor level means.

Definition. A linear combination of the factor level means has the form

$$L = \sum_{j=1}^a c_j \mu_j$$

where c_1, \dots, c_a are the coefficients that define the linear combination. A contrast is a linear combination of the factor level means that sum to zero, i.e., such that $\sum_{j=1}^a c_j = 0$.

Of course, in practice, we don't know the true population means, so the linear combination can be estimated such as

$$\hat{L} = \sum_{j=1}^a c_j \hat{\mu}_j$$

where $\hat{\mu}_j = \bar{y}_j$ are the estimated factor level means. Note that the variance of the linear combination has the form

$$\text{Var}(\hat{L}) = \sum_{j=1}^a c_j^2 \text{Var}(\hat{\mu}_j) = \sigma^2 \sum_{j=1}^a \frac{c_j^2}{n_j}$$

which is due to the fact that $\text{Var}(\hat{\mu}_j) = \sigma^2/n_j$ given that $\hat{\mu}_j = \bar{y}_j$ is a sample mean.

Suppose we want to test a null hypothesis of the form $H_0 : L = L_0$ versus the alternative hypothesis $H_1 : L \neq L_0$. We can use a t test statistic

$$T_0 = \frac{\hat{L} - L_0}{\sqrt{\widehat{\text{Var}}(\hat{L})}}$$

where $\widehat{\text{Var}}(\hat{L}) = \hat{\sigma}^2 \sum_{j=1}^a c_j^2/n_j$ uses the MSW to estimate σ^2 in the variance calculation. Assuming the null hypothesis $H_0 : L = L_0$ is true, the test statistic T_0 follows a Student's t distribution with $n - a$ degrees of freedom. So, if we only wanted to test a single linear combination, we could use a Student's t test. However, when there are $a > 2$ groups, we often want to test more than a single hypothesis.

5.2 Family-Wise Error Rate

Definition. When conducting $m \geq 2$ significance tests, the family-wise type I error rate (FWER) is the probability of making at least one type I error among the m tests.

As a reminder, for a single significance test, the type I error rate is defined as

$$\alpha = P(\text{Reject } H_0 \mid H_0 \text{ is true})$$

which is the probability of a false positive (i.e., rejecting a true null hypothesis). Suppose that we have a collection of m significance tests H_{0k} for $k = 1, \dots, m$, and assume that we use a significance level of α for each test. If the m tests are independent of one another, then the FWER can be written as

$$\text{FWER} = 1 - (1 - \alpha)^m$$

where $(1 - \alpha)^m$ is the probability of not making any type I error, which is simply the product of not making a type I error for each test. For example, with $m = 3$ independent tests, we have $\text{FWER} = 0.142625$, and with $m = 5$ independent tests, we have $\text{FWER} = 0.2262191$. However, in practice the m tests will not necessarily be independent of one another. When the tests are dependent on one another, there is no simple formula for calculating the FWER. This is because, when the tests are dependent, the FWER will depend on the dependence structure of the m tests—which is often unknown.

5.3 Bonferroni Correction

Suppose that we have m tests with an unknown dependence structure, and let α_k denote the type I error rate for the k -th test (for $k = 1, \dots, m$). In most cases, $\alpha_k = \alpha_0 \forall k$, where α_0 is the common type I error rate for each test. According to Boole's inequality, we know that

$$\text{FWER} \leq \sum_{k=1}^m P(\text{Reject } H_{0k} \mid H_{0k} \text{ is true}) = \sum_{k=1}^m \alpha_k$$

and note that this can be further simplified to $\text{FWER} = m\alpha_0$ if $\alpha_k = \alpha_0 \forall k$. Bonferroni's solution to the multiple comparison problem is to simply set the α_k significance levels such that $\sum_{k=1}^m \alpha_k = \alpha$, where α is the desired FWER. If $\alpha_k = \alpha_0 \forall k$, Bonferroni's correction simply sets $\alpha_0 = \alpha/m$, i.e., the significance level for each test is the desired FWER divided by the number of tests. The benefit of Bonferroni's method is that it works regardless of the dependence structure of the m tests. The downside is that Bonferroni's method can be overly-conservative when the m tests are independent. Note that for m independent tests, we have that $\text{FWER} = 1 - (1 - \alpha/m)^m \leq \alpha$ with the quality holding when $m = 1$.

5.4 Tukey's Honest Significant Difference

Tukey considered the problem of testing all of the $a(a-1)/2$ possible pairwise comparisons of the form $H_0 : \mu_j - \mu_k = 0 \forall j, k$ versus the alternative $H_1 : (\exists j, k)(\mu_j \neq \mu_k)$. Note that the estimated pairwise comparison $\hat{L}_{jk} = \hat{\mu}_j - \hat{\mu}_k$ is a contrast, given that $c_j = 1$ and $c_k = -1$ and $c_j + c_k = 0$. The variance of any pairwise comparison has the form

$$\text{Var}(\hat{L}_{jk}) = \sigma^2 \left(\frac{1}{n_j} + \frac{1}{n_k} \right)$$

which simplifies to $\text{Var}(\hat{L}_{jk}) = \sigma^2(2/n_0)$ if $n_j = n_k = n_0$. To test all pairwise comparisons, Tukey assumed that $n_j = n_k = n_0 \forall j, k$, and proposed using the test statistic

$$q = \frac{\sqrt{2}\hat{L}_{jk}}{\sqrt{\widehat{\text{Var}}(\hat{L}_{jk})}} = \frac{\bar{y}_j - \bar{y}_k}{\hat{\sigma}/\sqrt{n_0}}$$

Furthermore, Tukey showed that the test statistic q follows the *Studentized range distribution* with degrees of freedom $(a, n - a)$ if H_0 is true. Note that the Studentized range distribution is a family of distributions, similar to the F distribution, that is characterized by two degrees of freedom parameters. For balanced designs (i.e., when $n_j = n_k = n_0 \forall j, k$), using Tukey's procedure will exactly control the FWER across all $a(a-1)/2$ possible pairwise comparisons. When the design is unbalanced (i.e., when $n_j \neq n_k$ for some j, k), the procedure is known as the Tukey-Kramer method, and the FWER will be below (but not exactly equal to) the significance level α , i.e., the Tukey-Kramer method is conservative.

5.5 Scheffé's Method

Scheffé considered the problem of testing the significance of all possible contrasts of the factor level means. Specifically, Scheffé's procedure tests the null hypothesis $H_0 : \sum_{j=1}^a c_j \mu_j = 0$ for all coefficients satisfying $\sum_{j=1}^a c_j = 0$, versus the alternative hypothesis $H_1 : \sum_{j=1}^a c_j \mu_j \neq 0$ for at least one set of coefficients satisfying $\sum_{j=1}^a c_j = 0$. Note that all contrasts are equal to zero if and only if $\mu_j = \mu \forall j$. To prove this claim note that

- $\mu_j = \mu \forall j$ implies that $\sum_{j=1}^a c_j \mu_j = \mu \sum_{j=1}^a c_j = 0$ (this proves the “if” part)
- $\sum_{j=1}^a c_j \mu_j = 0$ for all contrasts means $\mu_j - \mu_k = 0 \forall j, k$ (this proves the “only if” part)

This implies that if we want to test all possible contrasts, we can simply use the omnibus F test that we previously discussed for testing the overall significance of the ANOVA model. Note that pairwise comparisons are one possible kind of contrast, but they are not the only kind of contrast—so Scheffé's procedure is testing more possibilities than Tukey's method.

5.6 Confidence Intervals

To form a $100(1 - \alpha)\%$ confidence interval for a generic linear combination $L = \sum_{j=1}^a c_j \mu_j$, we can use the general formula

$$\hat{L} \pm C_\alpha \sqrt{\widehat{\text{Var}}(\hat{L})}$$

where $\hat{L} = \sum_{j=1}^a c_j \hat{\mu}_j = \sum_{j=1}^a c_j \bar{y}_j$ is the estimated linear combination, $\widehat{\text{Var}}(\hat{L}) = \hat{\sigma}^2 \sum_{j=1}^a c_j^2 / n_j$ is the estimated variance of the linear combination, and C_α denotes some critical value that depends on the confidence level α . When forming a confidence interval for a single linear

combination, the critical value is defined as $C_\alpha = t_{n-a}^{(1-\alpha/2)}$, where $t_{n-a}^{(1-\alpha/2)}$ denotes the $(1-\alpha/2)$ quantile of Student's t distribution with $n-a$ degrees of freedom. However, when forming more than one confidence interval, the choice of the critical value C_α will depend on both the nature of the linear combinations (e.g., pairwise comparisons, contrasts, or other) and the number of linear combinations that are of interest:

For each possible method to control the FWER, the critical value would be defined as:

- None: $C_\alpha = t_{n-a}^{(1-\alpha/2)}$
- Bonferroni: $C_\alpha = t_{n-a}^{(1-\alpha/(2m))}$
- Tukey: $C_\alpha = q_{a,n-a}^{(1-\alpha)}/\sqrt{2}$
- Scheffé: $C_\alpha = \sqrt{(a-1)F_{a-1,n-a}^{(1-\alpha)}}$

where t_{n-a} denotes Student's t distribution, m denotes the number of confidence intervals (for Bonferroni), $q_{a,n-a}$ denotes the Studentized range distribution (for Tukey), and $F_{a-1,n-a}$ denotes the F distribution (for Scheffé). When the collection of linear combinations includes some linear combinations that are non-contrasts, i.e., where $\sum_{j=1}^a c_j \neq 0$, then Tukey's method and Scheffé's methods are not possible (so Bonferroni's method is the only option). If the collection of linear combinations only includes the $a(a-1)/2$ pairwise comparisons, then Tukey's method should be preferred because this method is exact. If the collection of linear combinations consists of only contrasts (some of which are not pairwise comparisons), then you could choose between Scheffé's and Bonferroni's method, and should prefer the method with the smaller C_α value.

Example 2. Revisiting the previous example, suppose that we want to form confidence intervals for each of the possible pairwise comparisons. In this case, we could choose between all three FWER correction methods: Bonferroni, Tukey, and Scheffé. Of course, we know that Tukey's method should be preferred for this situation, but we will still compare the three methods. Given that there are $n_0 = 10$ subjects in each condition, the variance of any pairwise comparison $L_{jk} = \mu_j - \mu_k$ is given by $\text{Var}(L_{jk}) = \sigma^2(2/n_0)$. Plugging in the mean squared error for the unknown σ^2 , we have that

$$\widehat{\text{Var}}(L_{jk}) = \hat{\sigma}^2(2/n_0) = (535.6/27)(2/10) = 1071.2/270 = 3.967407$$

For a 95% confidence interval, the different critical values that we could use are given by:

- None: $C_\alpha = t_{27}^{(0.975)} = 2.051831$
- Bonferroni: $C_\alpha = t_{27}^{(1-0.05/6)} = 2.552459$
- Tukey: $C_\alpha = q_{3,27}^{(0.95)} / \sqrt{2} = 2.479418$
- Scheffé: $C_\alpha = \sqrt{2F_{2,27}^{(0.95)}} = 2.590031$

Note that, of the four methods, the Tukey should be preferred because it produces the narrowest confidence interval that will control the FWER at the desired $\alpha = 0.05$ level across the three confidence intervals.

The three confidence intervals using Tukey's method are:

- $\hat{L}_{12} \pm (q_{3,27}^{(0.95)} / \sqrt{2}) \sqrt{\widehat{\text{Var}}(L_{jk})} = (27.4 - 21.2) \pm (2.479418) \sqrt{3.967407} = [1.2614; 11.1386]$
- $\hat{L}_{13} \pm (q_{3,27}^{(0.95)} / \sqrt{2}) \sqrt{\widehat{\text{Var}}(L_{jk})} = (21.8 - 21.2) \pm (2.479418) \sqrt{3.967407} = [-4.3386; 5.5386]$
- $\hat{L}_{23} \pm (q_{3,27}^{(0.95)} / \sqrt{2}) \sqrt{\widehat{\text{Var}}(L_{jk})} = (21.8 - 27.4) \pm (2.479418) \sqrt{3.967407} = [-10.5386; -0.6614]$

where $\bar{y}_1 = 21.2$ is the average number of words remembered in the fast condition, $\bar{y}_2 = 27.4$ is the average number of words remembered in the normal condition, and $\bar{y}_3 = 21.8$ is the average number of words remembered in the slow condition. In R, we could use

```
> TukeyHSD(amod)
Tukey multiple comparisons of means
95% family-wise confidence level

Fit: aov(formula = words ~ cond)

$cond
      diff      lwr      upr      p adj
normal-fast  6.2  1.261409 11.1385914 0.0117228
slow-fast    0.6 -4.338591  5.5385914 0.9513012
slow-normal -5.6 -10.538591 -0.6614086 0.0238550
```