

One-Way Analysis of Variance (ANOVA)

The basic scenario leading to an analysis of variance is one in which the goal is to explain the effect of one or more qualitative (categorical) variables upon a single quantitative variable. In these notes we assume a single qualitative independent variable, or a one-way analysis of variance.

Two equivalent models are commonly employed in one-way ANOVA: the cell means model, and the factor effects model. Although these notes primarily cover the cell means model, we will refer to the factor effects model where convenient and you are expected to be able to move between model descriptions.

Cell Means Model:

The following is meant to explain the notation used in the cell means model:

- Factor: Another name for the single qualitative independent variable.
- Level: one of the k distinct values of the factor involved in the analysis. Also called a *Treatment*.
- n_i , $1 \leq i \leq k$, is the size of the sample drawn from the i^{th} treatment population of the factor.
- $n = \sum n_i$ is the total sample size drawn from all treatments.
- μ_i : The mean value of the quantitative dependent (response) variable under the i^{th} treatment.
- Y_{ij} : The value of the response variable for the j^{th} observation from the i^{th} treatment in the sample.
- ε_{ij} : The difference between Y_{ij} and μ_i , i.e., $\varepsilon_{ij} = Y_{ij} - \mu_i$. Note: The assumption made in ANOVA is that the errors are independent and identically distributed as $N(0, \sigma^2)$.

For the j^{th} observation on the i^{th} treatment, the cell means model is given by $Y_{ij} = \mu_i + \varepsilon_{ij}$. Since the treatment mean μ_i is a constant, we have $E\{Y_{ij}\} = \mu_i$ and $\sigma_{Y_{ij}}^2 = \sigma^2$. Finally, this leads to the conclusion that, under the assumptions made about the distribution of the errors, the Y_{ij} are independent $N(\mu_i, \sigma^2)$.

Factor Effects Model:

The following is meant to explain the notation used in the factor effects model:

- μ : Under the most common interpretation, this is the unweighted average of all treatment means, i.e., $\mu = \frac{\sum \mu_i}{k}$. (In some situations, μ may be defined as a weighted average of the treatment means.)
- α_i : The effect of the i^{th} treatment. If μ is the unweighted mean of all treatment means, then $\sum \alpha_i = 0$.

For the j^{th} observation on the i^{th} treatment, the factor effects model is given by $Y_{ij} = \mu + \alpha_i + \varepsilon_{ij}$. Since the treatment mean $\mu + \alpha_i$ is a constant, we have $E\{Y_{ij}\} = \mu + \alpha_i$ and $\sigma_{Y_{ij}}^2 = \sigma^2$. Finally, this leads to the conclusion that, under the assumptions made about the distribution of the errors, the Y_{ij} are independent $N(\mu + \alpha_i, \sigma^2)$.

Fitting a One-Way ANOVA Model

The Analysis of Variance is a linear model because it can be expressed in the form $\mathbf{Y} = \mathbf{X}\beta + \boldsymbol{\varepsilon}$, where the forms of the design matrix \mathbf{X} and the coefficient vector β depend upon whether the cell means or factor effects models are being fitted. It can be shown that the least squares solution to the problem of fitting an ANOVA model to a sample is given by:

Cell Means Model: Under the least squares criterion, we seek to minimize $\sum_i \sum_j (Y_{ij} - \hat{\mu}_i)^2$, where $\hat{\mu}_i$ estimates

the i^{th} treatment mean μ_i . Rewritten as $\sum_i \sum_j (Y_{ij} - \hat{\mu}_i)^2 = \sum_{j=1}^{n_1} (Y_{1j} - \hat{\mu}_1)^2 + \sum_{j=1}^{n_2} (Y_{2j} - \hat{\mu}_2)^2 + \cdots + \sum_{j=1}^{n_k} (Y_{kj} - \hat{\mu}_k)^2$,

the only way to minimize the overall sum is to minimize each of the component sums. You will show in a homework problem that, in general, a sum of the form $\sum_j (x_j - b)^2$, where b is a constant, is minimized by

letting $b = \bar{x}$, where \bar{x} is the average of the x_j . Thus, the least squares estimates of the treatment means are the

corresponding sample means, $\hat{\mu}_i = \bar{Y}_i = \frac{1}{n_i} \sum_{j=1}^{n_i} Y_{ij}$. The following notation is used for the cell means model:

- \bar{Y}_i : The mean response for the sample drawn from the i^{th} treatment population. So, $\bar{Y}_i = \frac{1}{n_i} \sum_{j=1}^{n_i} Y_{ij}$
- $\bar{\bar{Y}}$: The mean response for the total sample, $\bar{\bar{Y}} = \frac{1}{n} \sum_{i=1}^k \sum_{j=1}^{n_i} Y_{ij}$, where $n = n_1 + n_2 + \cdots + n_k$
- e_{ij} : The residual for the j^{th} observation on the i^{th} treatment, $e_{ij} = Y_{ij} - \bar{Y}_i$

Factor Effects Model: For the factor effects model, under the least squares criterion, the estimates are the same as those obtained in fitting the cell means model, but the statistics being fitted differ slightly. In particular,

- $\hat{\alpha}_i$: The estimated effect of the i^{th} treatment on the mean of the response variable, given by $\hat{\alpha}_i = \bar{Y}_i - \bar{\bar{Y}}$.

Partitioning the Sums of Squares

In both the cell means and the factor effects model, we'll differentiate between the following deviations:

- $Y_{ij} - \bar{\bar{Y}}$ is the deviation of the j^{th} observation on the i^{th} treatment from the overall mean for the sample. (This corresponds to $Y_i - \bar{Y}$ in regression.)
- $\bar{Y}_i - \bar{\bar{Y}}$ is the deviation of the sample mean for the i^{th} treatment from the overall mean for the sample. (This corresponds to $\hat{Y}_i - \bar{Y}$ in regression.)
- $Y_{ij} - \bar{Y}_i$ is the deviation of the j^{th} observation on the i^{th} treatment from the sample mean for the treatment. (This corresponds to $Y_i - \hat{Y}_i$ in regression.)

The deviations above each lead to a sum of squares corresponding to a familiar sum of squares in regression.

- $SST = \sum_i \sum_j (Y_{ij} - \bar{Y})^2$ is the **T**otal **S**um of **S**quares in one-way ANOVA.
- $SSR = \sum_i n_i (\bar{Y}_i - \bar{Y})^2$ is the "Between Treatments" Sum of Squares. (Note: traditionally, this sum of squares goes by other names, such as the treatment sum of squares, for example. However, since there is no generally accepted choice of name, I've decided to use the name it would have in regression.)
- $SSE = \sum_i \sum_j (Y_{ij} - \bar{Y}_i)^2$ is the **E**rror **S**um of **S**quares, just as in regression. It's also called the "Within Treatments" sum of squares because it measures error from the treatment means.

If we accept for the moment that sums of squares for one-way ANOVA correspond exactly to their counterparts in regression, then it won't come as a surprise that the sums of squares add similarly: $SST = SSR + SSE$. We can also partition the degrees of freedom as follows:

- $df_{SST} = n - 1$, because we lose one of the n "bits" of information contained in the n observations when we must estimate the overall mean μ by \bar{Y} .
- $df_{SSR} = k - 1$, because we must estimate $k - 1$ treatment means μ_i after we've estimated the overall mean μ . (Note: the estimates of μ and of the k treatment means μ_i are *not* independent because \bar{Y} can be obtained from the k estimates \bar{Y}_i .)
- $df_{SSE} = n - k$

Note: the degrees of freedom partition in the same way as do the corresponding sums of squares:

$df_{SST} = df_{SSR} + df_{SSE}$. Although we haven't proved it, this is a result of the geometry of the vector space of observations, \mathbb{R}^n . The vector of deviations "Between Treatments" and the vector of deviations "Within Treatments" lie in orthogonal subspaces in \mathbb{R}^n . Thus, we obtain the same Pythagorean right triangle in \mathbb{R}^n seen in regression, where the "Between" and "Within" vectors form the legs, and the vector of "Total" deviations forms the hypotenuse. Thus, the sums of squares relation $SST = SSR + SSE$ is just a statement of the Pythagorean Theorem in \mathbb{R}^n , where SSR and SSE are the squared lengths of the "Between" and "Within" vectors, respectively. Finally, the degrees of freedom relation $df_{SST} = df_{SSR} + df_{SSE}$ is just a statement of the dimensionality of the subspaces where the vectors "live."

The table below summarizes the relationships between the sums of squares and their degrees of freedom for both a one-way analysis of variance with k treatments, and a multiple regression with $k-1$ independent variables.

Question: Why is a One-Way ANOVA with k levels equivalent to a multiple regression with p parameters (Betas)?

One-Factor ANOVA with k levels		Multiple Regression with p parameters	
Sums of Squares	df	Sums of Squares	df
SSR "Between"	$k-1$	SSR	$p-1$
SSE "Within"	$n-k$	SSE	$n-p$
SST "Total"	$n-1$	SST	$n-1$

Testing for a Differential Treatment Effect on the Response Variable

The hypothesis test for treatment effect has the following forms, both of which lead to the same ANOVA table:

Cell Means Model:

- ❖ $\mathbf{H}_0: \mu_1 = \mu_2 = \dots = \mu_k$. This is claim that all levels of the factor variable produce the same effect on the mean of the response variable.
- ❖ \mathbf{H}_A : At least two of the treatment means differ.

Factor Effects Model:

- ❖ $\mathbf{H}_0: \alpha_1 = \alpha_2 = \dots = \alpha_k = 0$. This is the same null hypothesis as the cell means model, but considering the differential treatment effects α_i instead of the treatment means μ_i .
- ❖ \mathbf{H}_A : At least two α_i are *not* zero. Equivalently, at least two of the treatment means differ.

All Roads (Really Do) Lead to F: the F - Ratio in One-Way ANOVA

Chi-Square Random Variables and the F Distribution

If the errors in a one-way analysis of variance model are independent $N(0, \sigma^2)$ random variables, then (under the null hypothesis of our test) $\frac{SSR}{\sigma^2} \sim \chi_{k-1}^2$, where χ_{k-1}^2 is the Chi-square distribution with $k - 1$ degree of freedom.

Similarly, $\frac{SSE}{\sigma^2} \sim \chi_{n-k}^2$.

Note: The degrees of freedom above agree with those resulting from a multiple regression model based on one qualitative variable with k levels. Remember, we would only create $k - 1$ dummy variables for the k distinct values of the factor. Thus, the $k - 1$ degrees of freedom for SSR reflects the $k - 1$ "slopes" estimated in the model

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_{k-1} X_{k-1} + \varepsilon.$$

Then the *F - Ratio*, $\frac{MSR}{MSE}$, that appears in the ANOVA table is the *ratio* of two independent chi-square

distributions divided by their respective degrees of freedom. Under the model assumptions, the *F - Ratio* follows an *F* distribution with degrees of freedom df_1 and df_2 , where df_1 and df_2 are the degrees of freedom of the chi-square variables in the numerator and denominator of the *F - Ratio*, respectively.

The F - Ratio for One-Way Analysis of Variance (the ANOVA Table)

The hypothesis for the test in one-way ANOVA takes the form \mathbf{H}_0 : All treatment means equal, versus, \mathbf{H}_A : Some treatment means differ. As with all hypothesis tests, the test is conducted under the assumption that the null hypothesis, \mathbf{H}_0 , is correct. Assuming the errors are iid $N(0, \sigma^2)$, the following are all true under the null hypothesis (Note: n is the total size of the sample):

- $\frac{SSR}{\sigma^2} \sim \chi_{k-1}^2$
- $\frac{SSE}{\sigma^2} \sim \chi_{n-k}^2$
- $\frac{MSR}{MSE} \sim F_{k-1, n-k}$

Under the null hypothesis (that all treatment means are equal), the following are also true:

- $E\{MSR\} = \sigma^2$
- $E\{MSE\} = \sigma^2$

So, under the null hypothesis of the test we expect the F - *Ratio* to be about 1. Clearly, an F - *Ratio* very different from one is evidence for the alternative hypothesis. The question is: is the test left-, right-, or two-tailed. The answer is: it's **right-tailed**. The reason is that under the alternative hypothesis, \mathbf{H}_A , the following are true: (Note: the α_i below are the *factor level* or **treatment** effects, i.e., the differential effects attributable to the treatments.)

- $E\{MSR\} = \sigma^2 + \frac{1}{k-1} \sum_i n_i \alpha_i^2$
- $E\{MSE\} = \sigma^2$

Notice that under \mathbf{H}_A the expected mean square for the model, $E\{MSR\}$, is greater than σ^2 , so the F - *Ratio* is expected to be greater than 1. Hence, we reject \mathbf{H}_0 in favor of \mathbf{H}_A , i.e., we conclude there is a differential treatment effect, for values of F significantly greater than 1. Although we could use a table of critical values of the F distribution (or better yet, a graphing calculator with an F distribution app) with $df_1 = k - 1$ and $df_2 = n - k$ to conduct the test at a fixed significance level α , we'll rely on the P -value in the ANOVA table of computer output.

Notice that $E\{MSR\} = \sigma^2 + \frac{1}{k-1} \sum_i n_i \alpha_i^2$ under the alternative hypothesis. So, we are more likely to be able to reject the null hypothesis (and conclude a treatment effect exists) if either of the following are true:

- At least one of the α_i^2 is large, i.e., the larger the treatment effect is, the more likely it is that we'll detect it!
- The n_i are large. The larger the samples are, the more precisely we can estimate the treatment effects α_i (making it easier to detect differences between them).

Analysis of Factor Level Effects

Suppose that the errors appear to be independent $N(0, \sigma^2)$ based upon an analysis of the residuals, and that the P -value for the model leads us to conclude that some of the treatment means differ. The next step is to investigate the nature of the differences in treatment effects. From your introductory statistics course, you may be familiar with the t -test for the difference in two population means, μ_1 and μ_2 . Although this is not the test

that is usually used in the analysis of variance, it is a good place to start a discussion about analyzing treatment effects in one-way ANOVA.

A *t*-Test for the Difference in the Means of Two Treatments

Suppose that our interest lay only in considering whether the mean value of the dependent variable is different for two different levels of the factor variable. If the assumption that the errors are independent $N(0, \sigma^2)$ appears reasonable, then we can construct confidence intervals, and conduct hypothesis tests, for the difference in the means, $\mu_2 - \mu_1$.

Confidence intervals and hypothesis tests are most straightforward when they involve a *single* parameter estimated by a *single* random variable! So, we should think of the difference in the means, $\mu_2 - \mu_1$, as a single parameter. Then the obvious choice for an estimator is the random variable $\bar{Y}_2 - \bar{Y}_1$, where \bar{Y}_1 and \bar{Y}_2 are the means of independent samples drawn from the two treatments. The following notation summarizes where we are at this point.

- n_1 and n_2 are the sizes of the samples drawn from treatment populations one and two, respectively.
- σ_1^2 and σ_2^2 , the variances for the dependent variable in each population, are assumed equal, i.e., $\sigma_1^2 = \sigma_2^2 = \sigma^2$, where σ^2 is simply the variance of the error.
- $\bar{Y}_1 = \frac{1}{n_1} \sum_{j=1}^{n_1} Y_{1j}$ and $\bar{Y}_2 = \frac{1}{n_2} \sum_{j=1}^{n_2} Y_{2j}$ are the sample means under the two treatments.

Now, any confidence interval for, or hypothesis test of, $\mu_2 - \mu_1$ will depend upon the distribution of the estimator $\bar{Y}_2 - \bar{Y}_1$. Proceeding step by step,

- $\bar{Y}_1 \sim N\left(\mu_1, \frac{\sigma^2}{n_1}\right)$, and $\bar{Y}_2 \sim N\left(\mu_2, \frac{\sigma^2}{n_2}\right)$
- $\bar{Y}_2 - \bar{Y}_1 \sim N\left(\mu_2 - \mu_1, \frac{\sigma^2}{n_1} + \frac{\sigma^2}{n_2}\right)$, because linear combinations of normal variables are also normal, and the variance of the difference of two *independent* random variables is the *sum* of their individual variances.

Although we've assumed constant variance σ^2 , we've made no assumption about the actual value of σ^2 . Therefore, we must estimate σ^2 from the samples taken from the populations. The most efficient way to estimate σ^2 is by through the *pooled* estimate $S_p^2 = \frac{(n_1 - 1)S_1^2 + (n_2 - 1)S_2^2}{n_1 + n_2 - 2}$, where,

- $S_1^2 = \frac{\sum_{j=1}^{n_1} (Y_{1j} - \bar{Y}_1)^2}{n_1 - 1}$, and $S_2^2 = \frac{\sum_{j=1}^{n_2} (Y_{2j} - \bar{Y}_2)^2}{n_2 - 1}$ are the usual sample variances for the two samples drawn.

Then, from well established theory, the statistic (just another name for a random variable computed from the data) $\frac{(\bar{Y}_2 - \bar{Y}_1) - (\mu_2 - \mu_1)}{S_p \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}$ has a t distribution with $df = n_1 + n_2 - 2$.

Example: Five measurements of the carbon content (in ppm) of silicon wafers were recorded on successive days of production. Can we conclude that the carbon content has changed?

Day 1	2.01	2.13	2.20	2.09	2.07
Day 2	2.31	2.41	2.23	2.19	2.26

(The values in the table below were calculated on my calculator)

Populations	Sample Size	Sample Mean	Sample Variance	Pooled Variance	df	Critical Value Used
Day 1	5	2.10	0.0050	0.0061	8	$t_{8,0.025} = 2.306$
Day 2	5	2.28	0.0072			

Note: Because the sample sizes were equal in this example, the pooled variance is just the average of the two sample variances. In general, however, the pooled variance is a *weighted* average of the sample variances, where greater weight is placed on the estimate derived from the larger sample. This should seem reasonable since larger samples tend to provide more accurate estimates, and therefore should carry more weight in the *pooled* estimate.

Then a 95% confidence interval for the difference in mean carbon content for the two days is given by

$$(2.28 - 2.10) \pm 2.306 \left(0.078 \sqrt{\frac{1}{5} + \frac{1}{5}} \right) = 0.18 \pm 0.114, \text{ or } \mathbf{(0.066 \text{ ppm}, 0.294 \text{ ppm})}.$$

Similarly, we can conduct a two-tailed test of the equality of the mean carbon content for the two days, with hypotheses

- ❖ $H_0: \mu_1 = \mu_2$
- ❖ $H_A: \mu_1 \neq \mu_2$

The P -value for this test of 0.00665 (obtained from my calculator) suggests that the mean carbon content of the wafers for the two days differs, a conclusion we could have reached based on the confidence interval for the difference derived above.

Now, you're probably wondering why I've spent so much time on a test that rarely gets used in the analysis of variance. The first reason is that t -tests of the difference between the means of two populations are popular in statistics, and this is a class in second semester statistics. The second reason is more important to the analysis of variance: why, if they *are* so popular, *don't* we use t -tests in ANOVA to evaluate the differences in treatment means? This answer involves the idea of the family-wise error rate.

The Family-Wise Error Rate

Suppose that a factor has k levels and that we've concluded, based on the P -value in the ANOVA table, that some level means are unequal. We then set out to investigate the differences with the goal of (hopefully) ranking the treatment means. If we decide to conduct t -tests for every possibly pairing of treatments, we discover that there

$$\text{are } \binom{k}{2} = \frac{k!}{2!(k-2)!} = \frac{k(k-1)}{2} \text{ ways to do this. The problem is not merely that this could involve conducting a}$$

large number of t -tests, but the much less obvious problem of the family-wise error rate for such a strategy.

In the two-tailed t -tests being considered here, the significance level of the test, α , is the probability of rejecting the statement $\mu_{\text{treatment A}} = \mu_{\text{treatment B}}$ when the means for the two treatments are equal, i.e., the probability of committing a **Type I** error. To repeat myself, $\alpha = P(\text{Concluding the means differ} \mid \text{the means are equal})$. But we are conducting not *one* such test, but $\binom{k}{2}$ such tests. The question is, what is the probability that *at least one* **Type I** error occurs for the *family* of $\binom{k}{2}$ tests. This is called the *family-wise error rate*, and as we'll see, it can be much larger than α .

If the $\binom{k}{2}$ tests were independent (they actually aren't because the same data is used in multiple tests), then it's easy to see that $P(\text{at least one Type I error}) = 1 - P(\text{no Type I errors}) = 1 - (1 - \alpha)^r$ where $r = \binom{k}{2}$ is the number of pairings. In fact, this forms an (approximate) lower bound on the family-wise rate, while an upper bound is given by $\min\{k\alpha, 1\}$. For $\alpha = 0.05$, the table below summarizes the family-wise error rates for a few choices of k .

Number of Levels, k	r	"Lower Bound" on $P(\text{at least one Type I error})$	Upper Bound on $P(\text{at least one Type I error})$
3	3	$1 - (0.95)^3 = \mathbf{0.14}$	$3 * 0.05 = \mathbf{0.15}$
4	6	$1 - (0.95)^6 = \mathbf{0.26}$	$6 * 0.05 = \mathbf{0.30}$
5	10	$1 - (0.95)^{10} = \mathbf{0.40}$	$10 * 0.05 = \mathbf{0.50}$
6	15	$1 - (0.95)^{15} = \mathbf{0.54}$	$15 * 0.05 = \mathbf{0.75}$

Alternatives to All Possible Pairwise t -Tests

The problem of family-wise error rates has attracted the attention of some of the biggest names in statistics, who have developed procedures for constructing simultaneous confidence intervals that can also be used to conduct pairwise tests of treatment means. They can be accessed in StatGraphics in the *Means Plot* window by clicking the right mouse button and selecting *Pane Options*.

- Fisher's Least Significant Difference (LSD) intervals: Named after Sir R. A. Fisher, this is the method that StatGraphics defaults to (it appears in some of my solutions for no better reason than this).
- Tukey's Honest Significant Difference (HSD) intervals: Named for John Tukey, who worked for AT&T back when it had more money than God (and better service), this method was specifically designed to control the family-wise Type I error rate for *all* possible pairwise comparisons of treatment means at a fixed α . In most cases, this is the set of intervals that are preferred.
- Scheffe Intervals: Named for Henry Scheffe, who, besides deriving his intervals, wrote a classic text on the analysis of variance. This procedure was about more than just confidence intervals and pairwise comparisons. It was designed for the related problem of drawing inference on *contrasts*. We haven't discussed contrasts yet, and we may not have time to this semester, but Scheffe came up with a way to conduct tests of all possible contrasts at a fixed family-wise rate α . Scheffe intervals, however, tend to be more conservative, i.e., wider, than HSD intervals.
- Bonferroni Intervals: One of the original attempts at solving the problem of family-wise error rates, Bonferroni intervals are still useful in certain situations. Generally, however, Tukey's HSD intervals are probably those most commonly employed to draw simultaneous inference in ANOVA at a fixed α .

Example: (This is the first example explored in the original notes on the analysis of variance.) As city manager, one of your responsibilities is purchasing. The city is looking to buy lightbulbs for the city's streetlights. Aware that some brands' lightbulbs might outlive other brands' lightbulbs, you decide to conduct an experiment. Seven lightbulbs each are purchased from four brands (GE, Dot, West, and a generic) and placed in streetlights. The lifetime of each of the 28 lightbulbs is then recorded in the file "**Lightbulbs**." Let's consider four different 95% confidence intervals for the difference of the means of the GE and Dot lightbulbs. (Actually, I'm only interested in the width of the intervals since they will all be centered about the same point estimate of the difference in the means.) The table below contains all of the relevant statistics for the four intervals to be created.

Brand	Sample Size	Sample Mean	Sample Variance		
GE	7	2.336	0.0460		
Dot	7	2.0	0.0213		
West	7	1.787	0.0152		
generic	7	2.1	0.0105		

1. First we will construct a simple **t**-interval based solely on the samples taken from the GE and Dot populations.

For $\mu_{GE} - \mu_{Dot}$, the pooled estimate of the variance is $s_p^2 = \frac{(7-1)s_{GE}^2 + (7-1)s_{Dot}^2}{7+7-2} = 0.03365$. The 95%

confidence interval for $\mu_{GE} - \mu_{Dot}$ is $(\bar{Y}_{GE} - \bar{Y}_{Dot}) \pm t_{7+7-2, 0.025} \sqrt{s_p^2 \left(\frac{1}{7} + \frac{1}{7} \right)} = (2.336 - 2.000) \pm$

$$t_{12, 0.025} \sqrt{0.03365 \left(\frac{1}{7} + \frac{1}{7} \right)} = 0.336 \pm 2.179(0.0981) = 0.336 \pm 0.214.$$

2. The **t**-interval for $\mu_{GE} - \mu_{Dot}$ using Fisher's LSD is similar to the interval above, but with the pooled estimate of the variance and the degrees of freedom derived from the *MSE* estimate of the error variance

σ^2 computed from the ANOVA table. The general form for a Fisher $(1-\alpha)100\%$ LSD confidence interval for

$\mu_{GE} - \mu_{Dot}$ is $(\bar{Y}_{GE} - \bar{Y}_{Dot}) \pm t_{n-k, \alpha/2} \sqrt{MSE \left(\frac{1}{7} + \frac{1}{7} \right)}$, where $n-k$ is the degrees of freedom associated the

error sum of squares *SSE*. For the light bulb data, these are $MSE = 0.02324$ and $df = 24$. The 95% Fisher *t*-

interval becomes $(2.336 - 2.000) \pm t_{24, 0.025} \sqrt{0.02324 \left(\frac{1}{7} + \frac{1}{7} \right)} = 0.336 \pm 2.064(0.0815) = 0.336 \pm 0.168$

3. The **t**-interval for $\mu_{GE} - \mu_{Dot}$ using Bonferroni's method is similar to the LSD interval above, but replacing α with α / r , where $r = \binom{k}{2}$ is the number of pairwise comparisons. A Bonferroni 95% **family-wise t**-interval

for $\mu_{GE} - \mu_{Dot}$ is given by $(2.336 - 2.000) \pm t_{24, 0.025/6} \sqrt{0.02324 \left(\frac{1}{7} + \frac{1}{7} \right)} = 0.336 \pm 2.875(0.0815) = 0.336 \pm 0.234.$

4. Tukey's HSD intervals use a critical value drawn from a **Studentized Range** distribution with $df_1 = k$ and $df_2 = n - k$ (compare this with the **F**-test in one-way analysis of variance where $df_1 = k - 1$ and $df_2 = n - k$). Tables for the Studentized Range distribution appear in statistics texts, or can be found online. The critical value for the Studentized Range is written $q_{k, n-k, \alpha}$. A Tukey 95% family-wise confidence interval for

$\mu_{GE} - \mu_{Dot}$ is given by $(2.336 - 2.000) \pm q_{4, 24, 0.05} \sqrt{\frac{0.02324}{2} \left(\frac{1}{7} + \frac{1}{7} \right)} = 0.336 \pm 3.90(0.0815) = 0.336 \pm 0.225$

- The Bonferroni interval is wider than Fisher's LSD, but has the advantage of guaranteeing a fixed level of confidence for the family of all pairwise comparisons. Unless you've decided to focus on one particular comparison before gathering the samples, i.e., a-priori, the Bonferroni is better because it guarantees a family-wise error rate.

- Tukey's interval was narrower than Bonferroni's. This will be the case whenever all pairwise comparisons are considered, as when the choice of comparisons is made *after* seeing the data (post-hoc). Tukey intervals are the most widely used when multiple pairwise comparisons are envisioned.

The results of this example are summarized below, where the margin of error for each type of interval is given.

Type of Interval Used	Margin of Error	Comments
t	0.214	Doesn't control the family-wise error rate at 0.05
Fisher's LSD	0.168	Narrower than t , but doesn't control the family-wise error rate at 0.05
Bonferroni	0.234	Controls the family-wise error rate at 0.05, but is conservative
Tukey's HSD	0.225	Controls the family-wise error rate for all pairwise comparisons at 0.05
Scheffe	<i>Not computed</i>	Controls the family-wise error rate for all possible <i>contrasts</i> at 0.05