

# Analysis of Variance

OPRE 6301

# Introduction . . .

---

The purpose of **analysis of variance** is to compare two or more populations of interval data. Specifically, we are interested in determining whether *differences* exist between the population *means*.

Although we are interested in a comparison of means, the method works by analyzing the sample *variance*. The intuitive reason for this is that by studying the “sources” of variation in the sample data, we can decide whether any observed differences among multiple sample means can be attributed to chance, or whether they are indicative of actual differences among the means of the corresponding populations.

Analysis of variance is one of the most widely used methods in statistics.

# One-Way Analysis of Variance . . .

---

Consider the following scenarios:

- We wish to decide on the basis of sample data whether there is really a difference in the effectiveness of three methods of teaching computer programming.
- We wish to compare the average yield per acre of wheat when different types of fertilizers are used.
- We wish to find out whether differences exist between the mean travel time from home to work/school along three different routes.
- We wish to find out whether differences exist between average starting salary for MBA graduates from different schools.

In all of these examples, we are trying to find out the effect of a *single* “factor” on a population; this explains the description “one-way.”

## Example: Advertising Strategy . . .

We will illustrate the method using the following concrete example.

An apple juice manufacturer is planning to develop a new product — a liquid concentrate.

The marketing manager has to decide how to market the new product. Three strategies are considered:

- Emphasize *convenience* of using the product.
- Emphasize the *quality* of the product.
- Emphasize the product's low *price*.

An experiment was conducted as follows:

- An advertising campaign was launched in three cities.
- In each city, only *one* of the three characteristics of the new product (convenience, quality, and price) was emphasized.

Weekly sales were recorded for twenty weeks following the beginning of the campaigns (see Xm15-01.xls).

	<b>Convnce</b>	<b>Quality</b>	<b>Price</b>
Weekly sales	529	804	672
	658	630	531
	793	774	443
	514	717	596
	663	679	602
	719	604	502
	711	620	659
Weekly sales	606	697	689
		706	675
		615	512
		492	691
	663	719	733
	604	787	698
	495	699	776
Weekly sales	485		561
	557		572
	353		469
	557	634	581
	542	580	679
	614	624	532

## Research Hypothesis

We wish to answer the following question: Do differences in sales exist between the test markets (i.e., advertising strategies)? In other words, our research hypothesis is that at least two of the three mean sales differ.

Therefore,

$$H_0 : \mu_1 = \mu_2 = \mu_3$$

$H_1$  : At least two of the  $\mu_j$ s differ

We now need to develop an appropriate statistic to test the above pair of hypotheses.

## Notation

Let

$k$  = total number of populations ( $k = 3$  in this example)

$n_j$  = sample size for the  $j$ th population ( $n_j = 20$  for all  $1 \leq j \leq k$  in this example)

$n = \sum_{j=1}^k n_j$ , the total number of observations

$X_{ij}$  = the  $i$ th observation from the  $j$ th population, where  $1 \leq j \leq k$  and for given  $j$ ,  $1 \leq i \leq n_j$ .

$\bar{X}_j$  = the sample mean of all observations from the  $j$ th population, for  $1 \leq j \leq k$ ; formally,

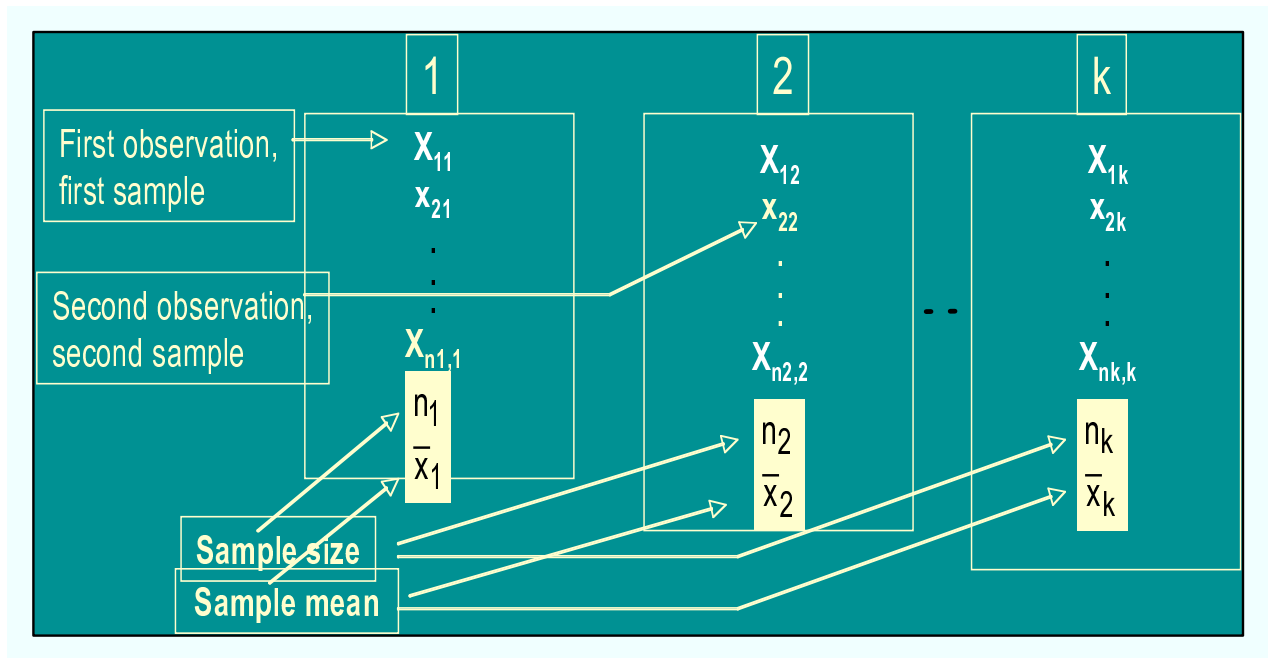
$$\bar{X}_j = \frac{1}{n_j} \sum_{i=1}^{n_j} X_{ij}.$$

$s_j$  = the standard deviation of all observations from the  $j$ th population

$\bar{\bar{X}}$  = the *grand* mean of all observations; formally,

$$\bar{\bar{X}} = \frac{1}{n} \sum_{j=1}^k \sum_{i=1}^{n_j} X_{ij} = \frac{1}{n} \sum_{j=1}^k n_j \bar{X}_j.$$

Thus,



Note that in general, the  $n_j$ s do *not* have to be the same.

## Generic Terminology

In the context of this example,

**Response Variable:** weekly sales, denoted by  $X$  above

**Responses:** actual sales values, which are the observed values of the  $X_{ij}$ s

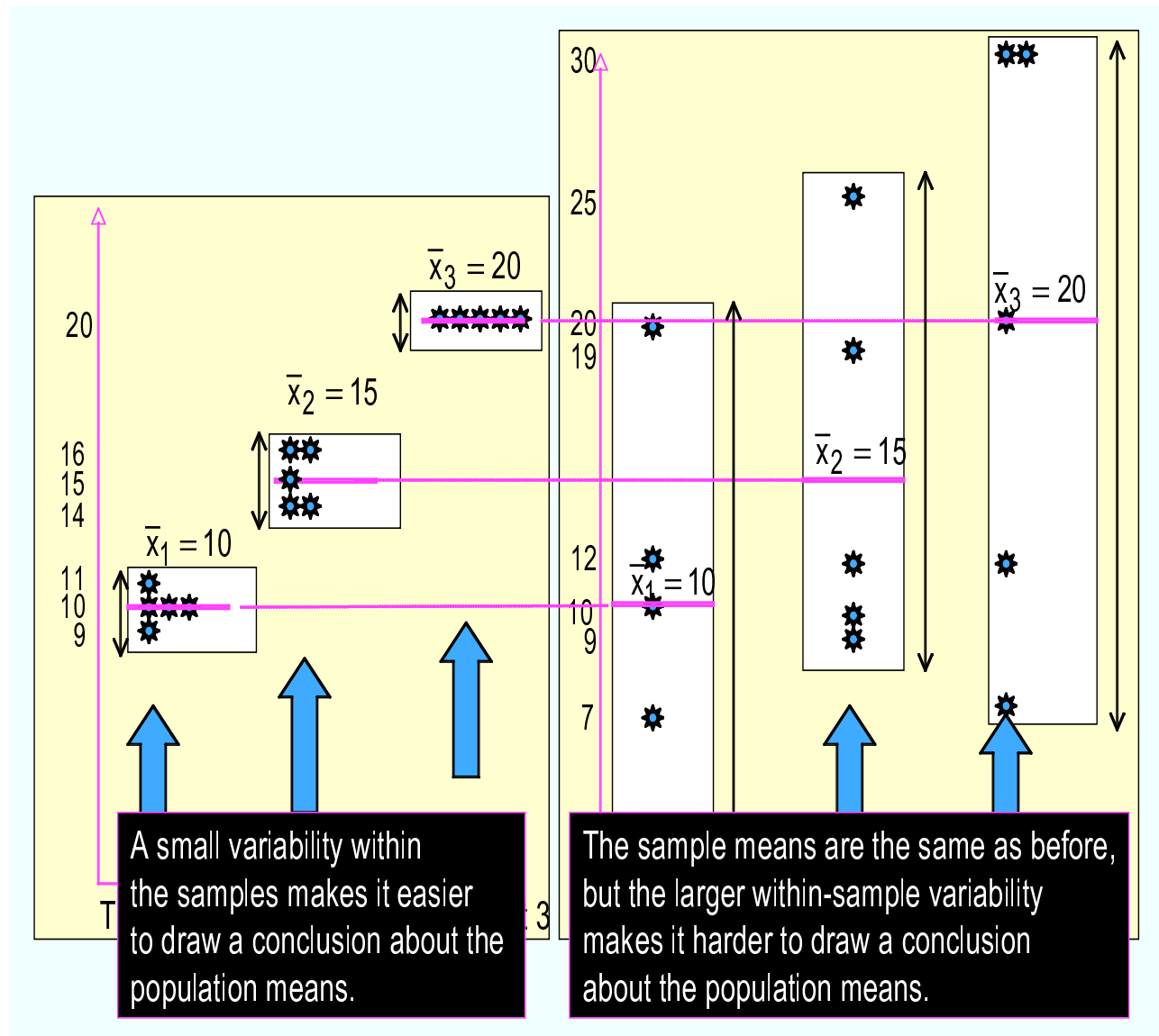
**Experimental Unit:** weeks in the three cities when we recored sales figures

**Factor:** the criterion by which we classify the populations — advertising strategy; another possible factor is advertising medium (TV, newspaper, or internet)

**Factor Levels:** possible **treatments** for a given factor — convenience, quality, or price

# Test Statistic

We will consider *two* measures of variability. The idea is depicted below ...



Formally, we assume that the  $X_{ij}$ s are independent and normally distributed with means  $\mu_j$  and a *common* variance  $\sigma^2$ . That is,

$$\bar{X}_{ij} = \mu_j + \epsilon_{ij}, \quad (1)$$

where the  $\epsilon_{ij}$ s, the **errors**, are normally distributed with mean 0 and variance  $\sigma^2$ .

Now, if the null hypothesis is true, say with all  $\mu_j$ s equal to  $\mu$ , then the key idea is that we can estimate  $\sigma^2$  in two ways ...

Method 1: For each  $j$ , the  $n_j$  observations  $X_{1j}, X_{2j}, \dots, X_{n_j,j}$  can be viewed as a sample of size  $n_j$  from a normal population with mean  $\mu$  and variance  $\sigma^2$ . It follows that

$$\frac{1}{\sigma^2} \sum_{i=1}^{n_j} (X_{ij} - \bar{X}_j)^2$$

has a  $\chi^2$  distribution with  $n_j - 1$  degrees of freedom (see equation (3) in notes for Chapter 12).

Furthermore, summing the above over  $j$  shows that

$$\frac{1}{\sigma^2} \sum_{j=1}^k \sum_{i=1}^{n_j} (X_{ij} - \bar{X}_j)^2$$

has a  $\chi^2$  distribution with  $n - k$  degrees of freedom (recall that  $\sum_{j=1}^k n_j = n$ ).

Since  $E(\chi^2) = \nu$ , we see that  $\sigma^2$  can be estimated by

$$\text{MSE} \equiv \frac{1}{n - k} \sum_{j=1}^k \sum_{i=1}^{n_j} (X_{ij} - \bar{X}_j)^2, \quad (2)$$

where MSE is for **mean square for errors** (cf. the white groupings in the previous figure).

Method 2: Recall that for each  $j$ , the sample mean  $\bar{X}_j$  also is a normally distributed variable with mean  $\mu$  and variance  $\sigma^2/n_j$ . Since we have one mean for each  $j$ , i.e., for each *treatment*, it is intuitive that one should be able to estimate  $\sigma^2$  by looking at the variability of the treatment means around the grand mean. Indeed, it can be shown that

$$\frac{1}{\sigma^2} \sum_{j=1}^k n_j (\bar{X}_j - \bar{\bar{X}})^2$$

(think of this as weighting the squared deviation  $(\bar{X}_j - \bar{\bar{X}})^2$  by  $n_j$ , the number of observations for treatment  $j$ ) has a  $\chi^2$  distribution with  $k - 1$  degrees of freedom.

Again, since  $E(\chi^2) = \nu$ , we see that  $\sigma^2$  can also be estimated by

$$\text{MST} \equiv \frac{1}{k - 1} \sum_{j=1}^k n_j (\bar{X}_j - \bar{\bar{X}})^2, \quad (3)$$

where MST is for **mean square for treatments** (cf. the horizontal pink lines in the previous figure).

Since both (2) and (3) are unbiased point estimates for  $\sigma^2$ , we expect the two estimates to be close. This suggests that we consider the statistic

$$F \equiv \frac{\text{MST}}{\text{MSE}}. \quad (4)$$

If the null hypothesis is true, we expect  $F$  to be close to 1. If, on the other hand, the null hypothesis is not true, we expect  $F$  to be pulled *up* by a greater MST (again, refer to the previous figure). Thus, the  $F$  statistic is a measure of *relative variability*.

How big an  $F$  value is considered sufficient for us to reject  $H_0$ ? The answer depends on the sampling distribution of the  $F$  statistic, which turns out to be the so-called (naturally)  $F$  distribution. (A discussion of the  $F$  distribution can be found on pp. 270–274 of the text.)

The  $F$  distribution is characterized by two parameters  $\nu_1$  and  $\nu_2$ . In our setting, these parameters correspond to the degrees of freedom for MST and MSE, respectively, and hence  $\nu_1 = k - 1$  and  $\nu_2 = n - k$ .

The  $F$  critical value for a given right-tail probability  $\alpha$  can be found by the Excel function  $\text{FINV}(\alpha, \nu_1, \nu_2)$ .

In our example, we have  $\nu_1 = 3 - 1 = 2$  and  $\nu_2 = 60 - 3 = 57$ . For  $\alpha = 0.05$ , this yields the critical value

$$\text{FINV}(0.05, 2, 57) = 3.16.$$

We are now ready to test  $H_0 \dots$

## The $F$ Test

The MST and the MSE are usually computed via

$$\text{MST} = \frac{\text{SST}}{k - 1}$$

and

$$\text{MSE} = \frac{\text{SSE}}{n - k},$$

respectively, where

$$\text{SST} \equiv \sum_{j=1}^k n_j (\bar{X}_j - \bar{\bar{X}})^2, \quad (5)$$

which is called the **sum of squares for treatment**, and

$$\text{SSE} \equiv \sum_{j=1}^k \sum_{i=1}^{n_j} (X_{ij} - \bar{X}_j)^2, \quad (6)$$

which is called the **sum of squares for errors**.

For our example, ...

SST:

$\bar{x}_1 = 577.55 \quad \bar{x}_2 = 653.00 \quad \bar{x}_3 = 608.65$

The grand mean is calculated by

$$\bar{\bar{X}} = \frac{n_1 \bar{x}_1 + n_2 \bar{x}_2 + \dots + n_k \bar{x}_k}{n_1 + n_2 + \dots + n_k}$$
$$\begin{aligned} \text{SST} &= \sum_{j=1}^k n_j (\bar{x}_j - \bar{\bar{X}})^2 \\ &= 20(577.55 - 613.07)^2 + \\ &\quad + 20(653.00 - 613.07)^2 + \\ &\quad + 20(608.65 - 613.07)^2 \\ &= 57,512.23 \end{aligned}$$

SSE:

$$s_1^2 = 10,775.00 \quad s_2^2 = 7,238.11 \quad s_3^2 = 8,670.24$$
$$\begin{aligned} \text{SSE} &= \sum_{j=1}^k \sum_{i=1}^{n_j} (X_{ij} - \bar{X}_j)^2 = (n_1 - 1)s_1^2 + (n_2 - 1)s_2^2 + (n_3 - 1)s_3^2 \\ &= (20 - 1)10,774.44 + (20 - 1)7,238.61 + (20 - 1)8,670.24 \\ &= 506,983.50 \end{aligned}$$

It follows that

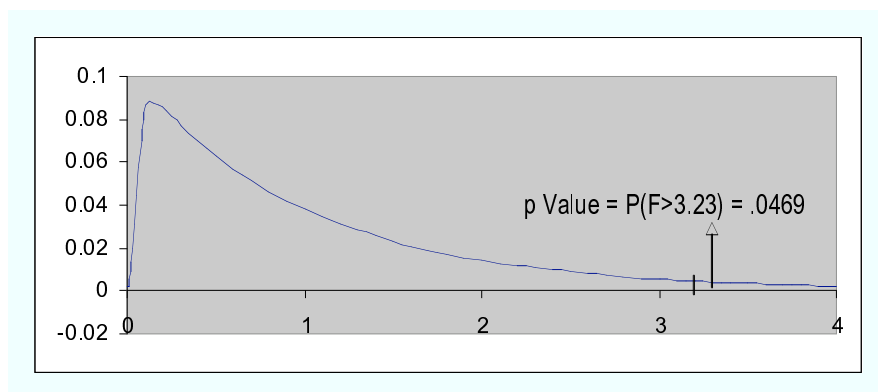
$$\begin{aligned} F &= \frac{SST/(k-1)}{SSE/(n-k)} \\ &= \frac{57512.23/2}{509983.50/57} \\ &= 3.23. \end{aligned}$$

Since 3.23 exceeds 3.16, the  $F$  critical value at  $\alpha = 0.05$ , there is sufficient evidence to reject  $H_0$  in favor of  $H_1$  and we conclude that at least one mean sales under these three strategies differ.

As usual, we can also determine the **right**-side  $p$ -value associated  $F = 3.23$  via

$$\text{FDIST}(3.23, 2, 57) = 0.0469,$$

which is seen to be lower than 0.05.



The Data Analysis tool “Anova: Single Factor” can be used to carry out  $F$  tests. This greatly simplifies our task. The output for our example is shown below.

Anova: Single Factor						
SUMMARY						
Groups	Count	Sum	Average	Variance		
Convenience	20	11551	577.55	10775.00		
Quality	20	13060	653.00	7238.11		
Price	20	12173	608.65	8670.24		
ANOVA						
Source of Variation	SS	df	MS	F	P-value	F crit
Between Groups	57512	2	28756	3.23	0.0468	3.16
Within Groups	506984	57	8894			
Total	564496	59				

Note that the last row of the output gives the total of SST and SSE, defined as:

$$SS \equiv \sum_{j=1}^k \sum_{i=1}^{n_j} (X_{ij} - \bar{X})^2. \quad (7)$$

It can be shown that

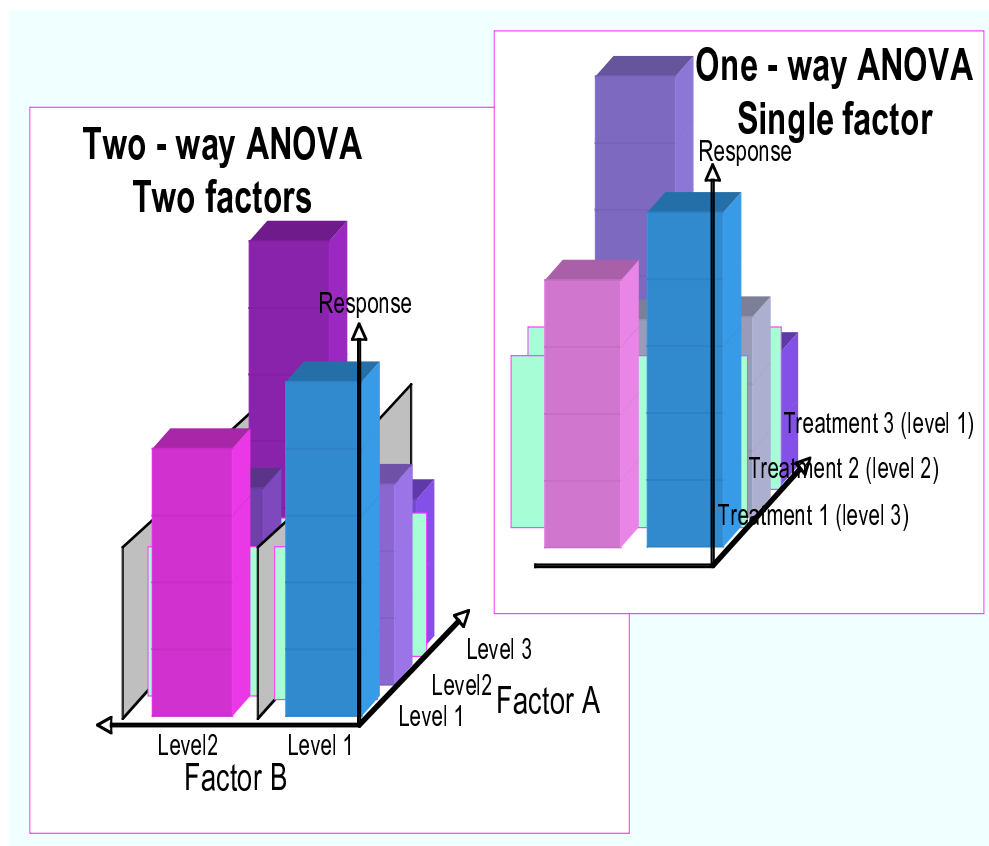
$$SS = SST + SSE \tag{8}$$

and that  $SS/\sigma^2$  follows the  $\chi^2$  distribution with  $n - 1$  degrees of freedom. The *decomposition* of SS into the sum of SST and SSE in (8) is helpful in visualizing how the  $F$  statistic is related to the two “sources” of variation, **between groups** and **within groups** (groups = treatments). We will encounter a similar decomposition in regression analysis.

# ANOVA: Two Factors ...

---

As noted earlier, advertising medium (TV, newspaper, or internet) could be considered as another factor that has an effect on sales. Visually, this means ...



The Data Analysis tool “Anova: Two-Factor ...” can be used for this. The idea is the same, and we leave out the details.

## Testing for Differences ...

---

After having rejected  $H_0$  in our example, a natural question arises: Which mean(s) is (are) different and furthermore, what is the rank order? A casual approach of course is to directly look at the  $\bar{X}_j$ s, but we should rely on a more rigorous statistical test.

Recall that for two populations having a common unknown variance, the confidence interval for the difference between the population means  $\mu_1 - \mu_2$  is given by

$$(\bar{X}_1 - \bar{X}_2) \pm t_{\alpha/2} \sqrt{s_p^2 \left( \frac{1}{n_1} + \frac{1}{n_2} \right)}, \quad (9)$$

where  $t_{\alpha/2}$  has  $n_1 + n_2 - 2$  degrees of freedom and

$$s_p^2 = \frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{n_1 + n_2 - 2}. \quad (10)$$

In fact, the  $F$  test with  $k = 2$  is equivalent to the  $t$  test; that is,  $F = t^2$ .

The idea then is to apply (9) and (10) to *all* pairs of sample means for treatments. However, we will make two revisions to (9) and (10) ...

Revision 1 (Fisher):

In this revision, we replace  $s_p^2$  in (10) by MSE. The reasoning is that MSE is more accurate, as it is based on data from *all* (not just two) treatment groups.

Formally, this results in the following decision rule: Reject  $\mu_i = \mu_j$  (for any pair of  $i$  and  $j$ ) if  $|\bar{X}_i - \bar{X}_j| > \text{LSD}_{ij}$ , where

$$\text{LSD}_{ij} = t_{\alpha/2} \sqrt{\text{MSE} \left( \frac{1}{n_i} + \frac{1}{n_j} \right)}, \quad (11)$$

with  $t_{\alpha/2}$  now having  $n - k$  degrees of freedom. The letters LSD stand for “Least Significant Difference.”

Revision 2 (Bonferroni):

Let  $\alpha_E$  (“ $E$ ” for experiment-wide) be the probability of committing *at least* one Type I error in *all*  $c = \binom{k}{2}$  pairwise tests. Then, under the assumption that the tests are only “mildly dependent,” we have the approximation:

$$\alpha_E \approx 1 - (1 - \alpha)^c. \quad (12)$$

Observe that as  $k$  grows, the above probability grows rapidly. It follows that Fisher’s procedure could easily result in an unacceptably high  $\alpha_E$ .

To remedy this problem, we will rely on the following (Bonferroni) bound:  $\alpha_E \leq c\alpha$  (the probability of a union is bounded above by the sum of the probabilities for all individual events). This bound implies that if we wish to ensure that the probability of committing a Type I error is no greater than  $\alpha_E$ , then we should set  $\alpha$  for the pairwise tests to  $\alpha_E/c$ .

Conclusion:  $t_{\alpha/2}$  in (11) should be replaced by  $t_{\alpha_E/(2c)}$ .

## Summary

After having rejected  $H_0$ , the following procedure can be used to conduct pairwise tests of differences in treatment means:

- Compute  $c$ , which is given by  $k(k - 1)/2$ .
- Set  $\alpha = \alpha_E/c$ , where  $\alpha_E$  is the targeted probability for making at least one Type I error in all pairwise tests.
- For a pair of  $i$  and  $j$ , conclude that  $\mu_i$  and  $\mu_j$  differ if

$$|\bar{X}_i - \bar{X}_j| > t_{\alpha/2} \sqrt{\text{MSE} \left( \frac{1}{n_i} + \frac{1}{n_j} \right)}, \quad (13)$$

where the degrees of freedom of  $t_{\alpha/2}$  is  $n - k$ .

## Example: Advertising Strategy — Continued

We failed to accept  $H_0$ . Which pair(s) of  $i$  and  $j$  has (have) different means?

Analysis: For  $k = 3$ , we have  $c = 3 \cdot 2/2 = 3$ . To achieve (say)  $\alpha_E = 0.05$ , we let  $\alpha = 0.05/3 = 0.0167$ . From Excel, we obtain

$$\text{TINV}(0.0167, 60 - 3) = 2.467.$$

Hence,

$ \bar{x}_1 - \bar{x}_2  =  577.55 - 653.0  = 75.45$	$t_{\alpha/2} \sqrt{MSE\left(\frac{1}{n_i} + \frac{1}{n_j}\right)} =$ $2.467 \sqrt{8894(1/20 + 1/20)} = 73.57$
$ \bar{x}_1 - \bar{x}_3  =  577.55 - 608.65  = 31.10$	
$ \bar{x}_2 - \bar{x}_3  =  653.0 - 608.65  = 44.35$	

We conclude that there is a significant difference between  $\mu_1$  and  $\mu_2$ .