

One-way Analysis of Variance

1. Beyond the unpaired t-test: $k=3$ or more independent samples

When one has two independent groups of observations, and a null hypothesis which states that they are random samples from 2 populations with identical means, it is customary to test that hypothesis using an independent-samples (or unpaired) t-test, as illustrated below:

$$\begin{aligned} H_0 : \mu_1 = \mu_2 & \quad \{ \text{null hypothesis} \} \\ H_{1a} : \mu_1 \neq \mu_2 & \quad \{ \text{non-directional alternative hypothesis} \} \end{aligned}$$

$$(1) \quad t = \frac{(\bar{X}_1 - \bar{X}_2) - (\mu_1 - \mu_2)}{S_{\bar{X}_1 - \bar{X}_2}} \quad \text{where}$$

$$(2) \quad S_{\bar{X}_1 - \bar{X}_2} = \sqrt{\left(\frac{s_{pooled}^2}{n_1} + \frac{s_{pooled}^2}{n_2} \right)}$$

= the **standard error of the difference** between 2 independent means

You may recall from an earlier chapter on z- and t-tests that:

$$s_{pooled}^2 = \text{pooled variance estimate} = \frac{SS_1 + SS_2}{n_1 + n_2 - 2} = \frac{SS_{\text{Within Groups}}}{df_{\text{Within Groups}}}$$

$$SS_1 = \sum (X - \bar{X})^2 \quad \text{for Group 1}$$

$$SS_2 = \sum (X - \bar{X})^2 \quad \text{for Group 2}$$

$$n_1 = \text{sample size for Group 1}$$

$$n_2 = \text{sample size for Group 2}$$

$$df = n_1 + n_2 - 2$$

But now suppose you had 4 independent samples, and the following null hypothesis:

$$H_0 : \mu_1 = \mu_2 = \mu_3 = \mu_4$$

How might you go about testing this null hypothesis? One possibility is to contrast each mean with every other mean using independent groups (unpaired) t-tests. For 4 groups, there would be 6 pairwise contrasts of means¹: 1-2, 1-3, 1-4, 2-3, 2-4, and 3-4. And if you set $\alpha = .05$ for each contrast, you would find that the overall probability of making at least 1 Type I error² is equal to approximately $6 \times .05 = .30$. (If you're wondering why I said approximately, all will be explained in a subsequent chapter on multiple comparison procedures.)

In this example, the .05 alpha for each contrast is called the **per-comparison** or **per-contrast** alpha³ (α_{PC}); and the overall alpha of .30 is called the **family-wise** alpha (α_{FW}).⁴ And note again that family-wise alpha is defined as the probability of making *at least one Type I error* while carrying out a family of comparisons.

Per-contrast alpha: the probability of making a Type I error for each individual contrast

Family-wise alpha: the probability of making at least one Type I error in a family of contrasts

To summarize, when you have 3 or more independent groups and a null hypothesis which states that all population means are equal, testing that null hypothesis by means of all pair-wise t-tests yields a family-wise alpha that is greater than desired—and the more groups you have, the greater the problem.

2. Solutions to the problem

2.1 Lowering the per-contrast alpha

It may occur to you that one could get around the problem of an inflated family-wise alpha by lowering the per-contrast alpha. If we were to set $\alpha_{PC} = .01$ for each of our 6 pairwise t-tests, for example, then the family-wise alpha would be approximately .06, which is quite close to the usual .05. I discuss this very technique, in fact, in a chapter on [multiple comparison procedures](#) (see the section on Fisher's LSD test). Suffice it to say for now that this is not the best solution when one is interested in testing the so-called **omnibus null hypothesis** ($H_0 : \mu_1 = \mu_2 = \dots = \mu_k$). Setting $\alpha_{PC} = .01$ results in too much loss of power. That is, it becomes too difficult to correctly reject a false null hypothesis.

¹ Note that we could have figured out how many pairwise contrasts there would be using the formula for combinations. $C_2^4 =$ the number of combinations of 4 means taken 2 at a time $= \frac{4!}{2!2!} = \frac{4(3)(2)(1)}{2(1)(2)(1)} = 6$ pairs of means.

² Type I error = rejection of the null hypothesis when it is true.

³ Some authors use the term "error rate" in place of alpha. I will tend to use alpha, because it is then clear that I am referring to Type I error rate.

⁴ You may also encounter the term "experiment-wise" error rate. Generally speaking, this is a broader term than family-wise: That is, an experiment may include 2 or more families of contrasts.

2.1 A better solution

As you have no doubt guessed, there is a better solution to the problem of inflated family-wise alpha in this situation. That solution is a technique called **AN**alysis **Of** **VAR**iance, or **ANOVA**. In this chapter, we will consider the simplest form of ANOVA, i.e., **one-way** (independent groups) ANOVA.

3. An overview of one-way ANOVA

Before we start looking at the details of the ANOVA calculations, I want to remind you that in general, hypothesis testing entails these three steps:

- 1) calculate a test statistic;
- 2) calculate a p-value—the probability of obtaining a test statistic as extreme or more extreme than the obtained value given a true null hypothesis;
- 3) decide whether you should act as if H_0 is true, or as if it is false.

Step 2 requires knowledge of the sampling distribution of the particular test statistic given a true null hypothesis. For a z -test, for example, the sampling distribution of the test statistic is the standard normal distribution. For a t -test, it is a t -distribution with the appropriate degrees of freedom. For the F -test you are about to encounter, it is one member of the family of F -distributions. The important thing to remember is that the same three steps are carried out for all of these tests. The tests differ only in the details of how we carry out the first two steps.

To *analyze* something is to break it down into components. In one-way *analysis of variance*, we partition the total variance of a set of scores into two components: A *between-groups* component, and a *within-groups* component. The ratio of these two components (*between-groups* component over *within-groups* component) is the test statistic, F . If the null hypothesis is true, then F will have an expected value of about 1, and a random sampling distribution that is described by one member of the family of F -distributions. If the null hypothesis is false, the expected value of F will be greater than 1, and the random sampling distribution of F will be shifted to the right.⁵ Using the appropriate F -distribution, we can calculate a p-value for our F -test—i.e., the probability of obtaining an F -value that large or larger, given that the null hypothesis is true. If the p-value is sufficiently low (usually .05 or less), we may reject the null hypothesis, and act as if it is false.

I will use the data set shown in Table 1 to illustrate the details of the analysis using two different conceptual approaches. These data are from a fictitious experimental investigation of the effect of distraction on students' ability to solve mental arithmetic problems. (Note that the number of subjects per group is deliberately low to make the calculations less time-consuming. In a real experiment, it would be advisable to use more subjects.) The independent variable is the amount of distraction participants experience present during the experiment (Low, Medium, or High), and the dependent variable is the number of problems solved in a fixed amount of time. The 3

⁵ As Howell (2007) points out, the expected value of F under a true null hypothesis is actually df_{error} divided by $df_{error} - 2$.

groups are of **equal size** ($n_1=n_2=n_3=5$; $N=15$). The null hypothesis states that these 3 groups are random samples from normally distributed populations with the same mean. It is further assumed, regardless of whether the null hypothesis is true or not, that the populations have equal variances. (This is known as the **homogeneity of variance** assumption.)

Table 1: Number of mental arithmetic problems solved.

	<u>Amount of Distraction</u>		
	Low	Medium	High
	5	5	3
	6	4	2
	2	2	6
	4	4	5
	2	5	4
Sum	19	20	20
N	5	5	5
Mean	3.8	4.0	4.0
SD	1.789	1.225	1.581
Variance	3.200	1.500	2.500
SS	12.8	6	10

Henceforth, I will refer to the Low, Medium, and High groups as groups 1, 2, and 3 respectively.

4. First conceptual approach to one-way ANOVA

Table 2 shows the same data we saw in Table 1, but in a way that will be more convenient for illustrating this conceptual approach to ANOVA. There is one row for each participant in the experiment. Each subject's group number is coded in the second column, and the dependent variable (number of problems solved) is shown in the Y-column.

The grand mean of all Y-scores is shown in the column headed with \bar{Y}_{\cdot} . (The *dot* subscript indicates *over all scores*.) The group mean appropriate for each score is given in the column headed by \bar{Y}_j (i.e., j can range from 1 to 3).

The final 3 columns in Table 2 give three different *deviation scores*:

- The deviations of raw scores from the grand mean of all the scores: $(Y_i - \bar{Y}_{\cdot})$
- The deviations of group means from the grand mean: $(\bar{Y}_j - \bar{Y}_{\cdot})$
- The deviations of raw scores from group means: $(Y_i - \bar{Y}_j)$

Subscripts. The i subscript is used to index the various subjects. In this case, it can range from 1 to 15. The j subscript is an index for group. In general, it ranges from 1 to k , where k is the number of groups. So in this instance, $k = 3$. The *dot* subscript indicates *over all scores*.

Table 2: Number of mental arithmetic problems solved (version 2).

Subject	Group	Y	\bar{Y}_{\cdot}	\bar{Y}_j	$(Y - \bar{Y}_{\cdot})$	$(\bar{Y}_j - \bar{Y}_{\cdot})$	$(Y - \bar{Y}_j)$
1	1	5	3.933	3.800	1.067	-0.133	1.200
2	1	6	3.933	3.800	2.067	-0.133	2.200
3	1	2	3.933	3.800	-1.933	-0.133	-1.800
4	1	4	3.933	3.800	0.067	-0.133	0.200
5	1	2	3.933	3.800	-1.933	-0.133	-1.800
6	2	5	3.933	4.000	1.067	0.067	1.000
7	2	4	3.933	4.000	0.067	0.067	0.000
8	2	2	3.933	4.000	-1.933	0.067	-2.000
9	2	4	3.933	4.000	0.067	0.067	0.000
10	2	5	3.933	4.000	1.067	0.067	1.000
11	3	3	3.933	4.000	-0.933	0.067	-1.000
12	3	2	3.933	4.000	-1.933	0.067	-2.000
13	3	6	3.933	4.000	2.067	0.067	2.000
14	3	5	3.933	4.000	1.067	0.067	1.000
15	3	4	3.933	4.000	0.067	0.067	0.000
Sums:					0.000	0.000	0.000

Notice that for each person, the sum of the final two deviations is equal to the first deviation. That is, the deviation of their *raw score from the grand mean* is equal to the deviation of their *group mean from the grand mean* plus the deviation of their *raw score from their group mean*. Putting it in symbols:

$$(Y_i - \bar{Y}_{\cdot}) = (\bar{Y}_j - \bar{Y}_{\cdot}) + (Y_i - \bar{Y}_j) \quad (3)$$

But notice as well that the sum of the deviation scores in each of the final 3 columns is zero.

You may recall that by definition, the mean of a distribution is the exact balancing point of the distribution. The sum of the deviations to the left of the mean is exactly equal to the sum of the deviations to the right of the mean. Putting it another way, the sum of all of the deviations about the mean (both positive and negative) is always equal to zero. It is not surprising, therefore, that the column sums shown in Table 2 are all equal to zero, because the final 3 columns all show deviations about various means.

Because the deviations about a mean always sum to zero, the deviation scores are frequently *squared* before summing (e.g., when calculating a variance or a SD). By definition, the sum of the squared deviations about the mean is a minimum.

4.1 Partitioning the total sum of squares

Let us look at the Table 2 data again, but with *squared deviations* in place of deviations in the final 3 columns.

Table 3: Partitioning of the Total Sum of Squares in One-way ANOVA.

Subject	Group	Y	\bar{Y}_{\cdot}	\bar{Y}_j	$(Y - \bar{Y}_{\cdot})^2$	$(\bar{Y}_j - \bar{Y}_{\cdot})^2$	$(Y - \bar{Y}_j)^2$
1	1	5	3.933	3.800	1.138	0.018	1.440
2	1	6	3.933	3.800	4.271	0.018	4.840
3	1	2	3.933	3.800	3.738	0.018	3.240
4	1	4	3.933	3.800	0.004	0.018	0.040
5	1	2	3.933	3.800	3.738	0.018	3.240
6	2	5	3.933	4.000	1.138	0.004	1.000
7	2	4	3.933	4.000	0.004	0.004	0.000
8	2	2	3.933	4.000	3.738	0.004	4.000
9	2	4	3.933	4.000	0.004	0.004	0.000
10	2	5	3.933	4.000	1.138	0.004	1.000
11	3	3	3.933	4.000	0.871	0.004	1.000
12	3	2	3.933	4.000	3.738	0.004	4.000
13	3	6	3.933	4.000	4.271	0.004	4.000
14	3	5	3.933	4.000	1.138	0.004	1.000
15	3	4	3.933	4.000	0.004	0.004	0.000
Sums:					28.933	0.133	28.800

The Table 3 column headed $(Y - \bar{Y}_{\cdot})^2$ shows the *squared deviation of each score from the grand mean*. The total at the bottom of this column is the *sum of the squared deviations about the grand mean*, or SS_{Total} . (If you were to divide SS_{Total} by the total N minus 1, you would have the variance of all the Y -scores.)

The next column, headed $(\bar{Y}_j - \bar{Y}_{\cdot})^2$, shows the *squared deviation of each score's Group Mean from the Grand Mean*. The sum of these squared deviation scores is equal to the *between-groups sum of squares*, or $SS_{Between-groups}$. (You may also see it referred to as the *treatment sum of squares*, or $SS_{Treatment}$.)

The final column, headed $(Y - \bar{Y}_j)^2$, shows the *squared deviation of each score from its own group mean*. The sum of these squared deviation scores is equal to the *within-groups sum of squares*, or $SS_{Within-groups}$. Another way to think of it is that the sum of the squared deviations for the first 5 rows (the green cells) is SS_1 , the SS for group 1; the sum of the squared deviations for the next 5 rows (the yellow cells) is SS_2 , the SS for group 2; and the sum of the squared deviations for the last 5 rows (the blue cells) is SS_3 , the SS for group 3. So, $SS_{Within-groups} = SS_1 + SS_2 + SS_3$ for this example. More generally, $SS_{Within-groups} = SS_1 + SS_2 + \dots + SS_k$.

In Table 2, we saw that the deviation scores were additive within each row. That is, $(Y_i - \bar{Y}_\bullet) = (\bar{Y}_j - \bar{Y}_\bullet) + (Y_i - \bar{Y}_j)$ for each subject. However, squaring of the deviations destroys that additivity within rows. That is to say,

$$(Y - \bar{Y}_\bullet)^2 \neq (\bar{Y}_j - \bar{Y}_\bullet)^2 + (Y - \bar{Y}_j)^2 \quad (4)$$

But all is not lost. Notice that if we sum across all subjects, the additivity is restored—thank heavens!

$$\begin{array}{rcccc} \sum (Y_i - \bar{Y}_\bullet)^2 & = & \sum (\bar{Y}_j - \bar{Y}_\bullet)^2 & + & \sum (Y_i - \bar{Y}_j)^2 \\ \downarrow & & \downarrow & & \downarrow \\ 28.933 & = & 0.133 & + & 28.800 \end{array} \quad (5)$$

If we now change to more conventional ANOVA terminology, we can restate equation (5) as follows:

$$SS_{\text{Total}} = SS_{\text{Between-groups}} + SS_{\text{Within-groups}} \quad (6)$$

4.2 Partitioning the total degrees of freedom

To this point, we have taken the *total sum of squares* for our set of scores, and partitioned it into a *between-groups sum of squares*, and a *within-groups sum of squares*. But as indicate in the general overview of one-way ANOVA, what we really need to have is two estimates of the *variance* of the population from which we have sampled our scores: a *between-groups* estimate, and a *within-groups* estimate.

As you may recall, one obtains a variance estimate by dividing a sum of squares by its degrees of freedom. In the case of the good old sample variance, for example:

$$s_y^2 = \frac{SS_y}{df_y} = \frac{\sum (Y - \bar{Y})^2}{n - 1} \quad (7)$$

Therefore, our next step is to partition df_{Total} , the total degrees of freedom for our set of scores, into $df_{\text{Between-groups}}$ and $df_{\text{Within-groups}}$.

As you might expect, the *total degrees of freedom* is equal to the total number of scores minus one. In symbols,

$$df_{\text{Total}} = N - 1 \quad \text{where} \quad N = \sum_{j=1}^k n_j = \text{Total number of scores} \quad (8)$$

It may not surprise you to learn that the *between-groups degrees of freedom* is equal to the number of groups minus one, or $k-1$:

$$df_{\text{Between-groups}} = k - 1 \quad \text{where } k = \text{number of groups} \quad (9)$$

Finally, the within-groups degrees of freedom is equal to the degrees of freedom within group 1 plus the degrees of freedom within group 2, and so on up to group k . The degrees of freedom within any group is equal to the number of scores within that group minus one (i.e., $df_j = n_j - 1$). So for our example,

$$df_{\text{Within-groups}} = df_1 + df_2 + df_3 = (n_1 - 1) + (n_2 - 1) + (n_3 - 1) \quad (10)$$

And in general,

$$df_{\text{Within-groups}} = \sum_{j=1}^k df_j \quad \text{where } df_j = n_j - 1 \quad (11)$$

An alternative, and perhaps simpler expression for the *within-groups df* is:

$$df_{\text{Within-groups}} = N - k \quad (12)$$

Plugging in the numbers for our data set, we get the following:

$$\begin{aligned} df_{\text{Total}} &= N - 1 = 14 \\ df_{\text{Between-groups}} &= k - 1 = 2 \\ df_{\text{Within-groups}} &= N - k = 12 \end{aligned}$$

It is no coincidence that $df_{\text{Total}} = df_{\text{Between-groups}} + df_{\text{Within-groups}}$. This will always be the case for any data set. So, both the *sums of squares* and the *degrees of freedom* are additive in one-way ANOVA. This additivity is often illustrated using a *partitioning diagram*, as in Figure 1.

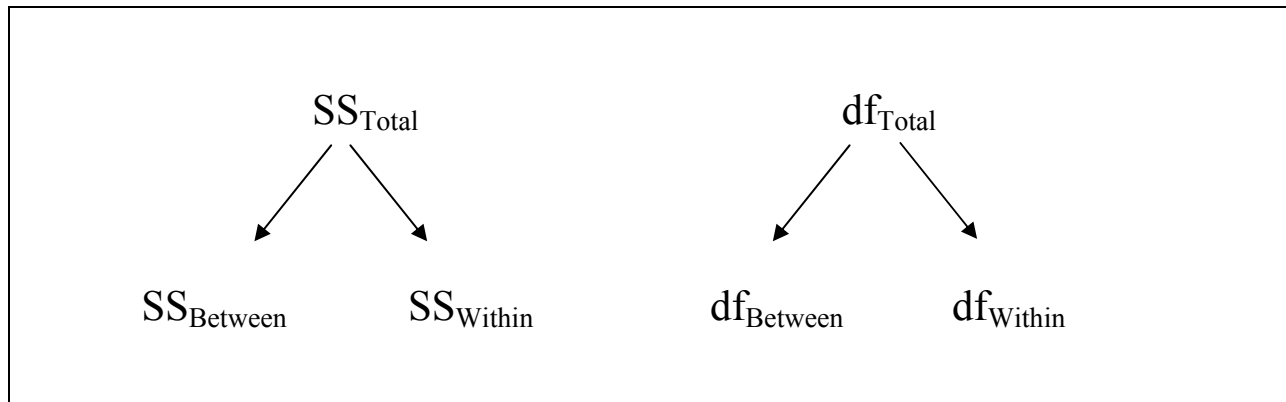


Figure 1: Partitioning diagram for one-way ANOVA.

4.3 Computing the two variance estimates

By taking each sum of squares (between- and within-groups) and dividing by the corresponding degrees of freedom, we obtain two variance estimates we need:

$$s_{\text{Between-groups}}^2 = \frac{SS_{\text{Between-groups}}}{df_{\text{Between-groups}}} \quad (13)$$

$$s_{\text{Within-groups}}^2 = \frac{SS_{\text{Within-groups}}}{df_{\text{Within-groups}}} \quad (14)$$

In the context of ANOVA (and linear regression), it is conventional to refer to these variance estimates as Mean Squares—or *MS* for short. So the last two equations would normally be written as follows:

$$MS_{\text{Between}} = \frac{SS_{\text{Between-groups}}}{df_{\text{Between-groups}}} = \frac{0.133}{2} = 0.067 \quad (15)$$

$$MS_{\text{Within}} = \frac{SS_{\text{Within-groups}}}{df_{\text{Within-groups}}} = \frac{28.800}{12} = 2.400 \quad (16)$$

Why variance estimates are called Mean Squares. A population variance is equal to the sum of the squared deviations about the mean divided by N : $\sigma_Y^2 = \sum (Y - \bar{Y})^2 / N$. So the population variance is really the mean of the squared deviations about the mean, or the **mean square(d) deviation** about the mean. This is why the term *mean square* is used in place of *variance*.

4.4 Comparing the two variance estimates

As mentioned earlier, F , the test statistic we compute for ANOVA, is equal to the ratio of these two variance estimates:

$$F_{(k-1, N-k)} = \frac{MS_{\text{Between}}}{MS_{\text{Within}}}, \quad \text{so for our example,} \quad (17)$$

$$F_{(2,12)} = \frac{0.067}{2.400} = 0.028$$

The null hypothesis we are testing states that our k groups are all random samples from populations with identical means (and variances). For our data set, the null hypothesis states that $\mu_1 = \mu_2 = \mu_3$.

If the null hypothesis is true, the random sampling distribution of the F -value we calculate will have an F -distribution with degrees of freedom equal to $k-1$ (in the numerator) and $N-k$ (in the denominator). As I mentioned earlier, there is a family of F -distributions, just as there is a family of t -distributions. F -distributions have a minimum value of 0, and tend to be positively skewed (i.e., they have a long tail to the right). Three members of the family of F -distributions are shown in Figure 2.

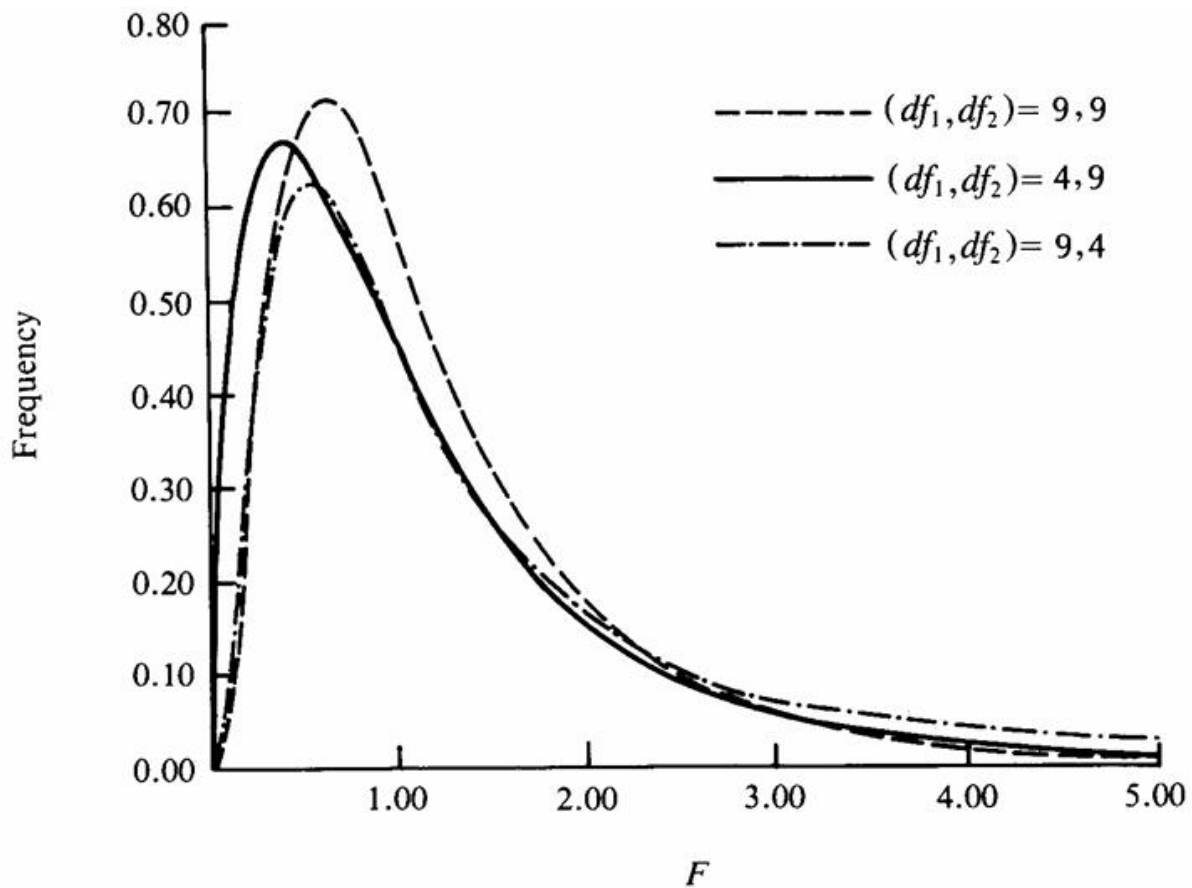


Figure 2: Examples of F -distributions.

So we may use the appropriate F -distribution to obtain either a critical value of F , or a p -value for our F -test. The p -value for an F -test is equal to the *area under the curve* to the right of the F -value you computed. This area may be obtained using integral calculus, which is not all that difficult nowadays. But in years gone by, it was considerably more difficult. For that reason, the common approach used to be looking up a critical value of F in a table, and rejecting H_0 if your computed F -value was equal to or greater than the critical value. But now, the preferred

approach is to report the actual p -value, and reject H_0 if p is less than some predetermined value (often 0.05).

All of the major statistical software packages include the p -value in the output. There are also good standalone programs one can use to compute p -values for common statistical tests. For example, you can download *StaTable* a free program from www.cytel.com. Alternatively, you can use the CDF.F⁶ function in SPSS as follows:

```
COMPUTE p = 1 - CDF.F(f, df1, df2) .
EXECUTE .
format p (f5.3) .
```

I used the CDF.F function in SPSS to compute $p = 0.972$ for our F -test. This means that if the null hypothesis is true, the probability of obtaining an F -value equal to or greater than 0.028 is 0.972. Therefore, we do not have sufficient evidence to allow rejection of H_0 . (Hey, let's face it—we weren't even close in this case!)

5. The ANOVA Summary Table

The results of a one-way ANOVA are often summarized in a table. Table 4 summarizes the results of our analysis.

Table 4: ANOVA summary for data in Table 1.

Source of Variation	SS	df	MS	F	p
Between Groups	0.133	2	0.067	0.028	0.972
Within Groups	28.800	12	2.400		
Total	28.933	14			

The p -value in the final column gives us the probability of obtaining an F -ratio of 0.028 or greater, given that the null hypothesis is true (and the assumptions of the analysis are met). If this conditional probability value is quite small (i.e., $p \leq .05$ by convention), we may reject the null hypothesis, and conclude that there is a statistically significant treatment effect.

⁶ The CDF in CDF.F stands for cumulative distribution function. It returns the proportion of area under the F -distribution (with the given degrees of freedom) to the left of the particular F -value you enter.

6. Another example (using the 1st conceptual approach)

In the preceding analysis, we found no evidence to suggest that increasing the amount of distraction (at least in the range examined by the experimenter) impairs students' ability to solve arithmetic problems. But consider the data in Table 5. These data come from a similar (fictitious) experiment that used **more difficult** arithmetic problems. Once again, the dependent variable (Y) is the number of problems solved in a given amount of time.

Table 5: Partitioning SS_{Total} for an experiment with more difficult arithmetic problems.

Subject	Group	Y	\bar{Y}_{\bullet}	\bar{Y}_j	$(Y - \bar{Y}_{\bullet})^2$	$(\bar{Y}_j - \bar{Y}_{\bullet})^2$	$(Y - \bar{Y}_j)^2$
1	1	5	6.867	4.800	3.484	4.271	0.040
2	1	6	6.867	4.800	0.751	4.271	1.440
3	1	2	6.867	4.800	23.684	4.271	7.840
4	1	5	6.867	4.800	3.484	4.271	0.040
5	1	6	6.867	4.800	0.751	4.271	1.440
6	2	9	6.867	5.400	4.551	2.151	12.960
7	2	5	6.867	5.400	3.484	2.151	0.160
8	2	3	6.867	5.400	14.951	2.151	5.760
9	2	7	6.867	5.400	0.018	2.151	2.560
10	2	3	6.867	5.400	14.951	2.151	5.760
11	3	7	6.867	10.400	0.018	12.484	11.560
12	3	10	6.867	10.400	9.818	12.484	0.160
13	3	11	6.867	10.400	17.084	12.484	0.360
14	3	12	6.867	10.400	26.351	12.484	2.560
15	3	12	6.867	10.400	26.351	12.484	2.560
Sums:					149.733	94.533	55.200

The means and standard deviations for the 3 groups in this experiment are as follows:

$$\bar{X}_1 = 4.80 \quad (SD_1 = 1.64)$$

$$\bar{X}_2 = 5.40 \quad (SD_2 = 2.61)$$

$$\bar{X}_3 = 10.40 \quad (SD_3 = 2.07)$$

As we saw in the previous example, the sums at the bottom of the final 3 columns in Table 5 are the SS_{Total} , SS_{Between} , and SS_{Within} respectively. We can plug these into the formulae we saw earlier:

$$(18) \quad MS_{\text{Between}} = \frac{SS_{\text{Between}}}{df_{\text{Between}}} = \frac{SS_{\text{Between}}}{k-1} = \frac{94.533}{2} = 47.267$$

$$(19) \quad MS_{\text{Within}} = \frac{SS_{\text{Within}}}{df_{\text{Within}}} = \frac{SS_{\text{Within}}}{N_{\text{Total}} - k} = \frac{55.2}{12} = 4.6$$

$$(20) \quad F_{(2,12)} = \frac{MS_{\text{Between}}}{MS_{\text{Within}}} = \frac{47.267}{4.600} = 10.275$$

The results of this analysis are summarized in Table 6. Note that the p -value for this F -test is equal to 0.003, which is well below the conventional alpha level of 0.05. Therefore, we may reject the null hypothesis, and conclude that not all population means are identical. Putting it another way, we may conclude that the time required to solve the more difficult arithmetic problems does depend on the amount of distraction.

Table 6: ANOVA summary for data in Table 5 (more difficult problems).

Source of Variation	SS	df	MS	F	P
Between Groups	94.533	2	47.267	10.275	0.003
Within Groups	55.200	12	4.600		
Total	149.733	14			

7. A second conceptual approach to one-way ANOVA

This second conceptual approach to one-way ANOVA is based on the **central limit theorem (CLT)**. You may have encountered the CLT in a chapter on z - and t -tests. It tells us, among other things, that:

- The mean of the sampling distribution of \bar{X} = the population mean ($\mu_{\bar{X}} = \mu_X$)
- The SD of the sampling distribution of \bar{X} = the standard error (SE) of the mean = the population standard deviation divided by the square root of the sample size
($\sigma_{\bar{X}} = \sigma_X / \sqrt{n}$)

7.1 The between-groups variance estimate

When contemplating the between-groups sum of squares, I find it useful to temporarily disregard the fact that the group means are means, and just treat them as a sample of raw scores. Doing so with the group means in Table 1, we have the following set scores:

$$X_1 = 3.8 \quad X_2 = 4.0 \quad X_3 = 4.0$$

The mean and variance of these scores can be calculated as follows:

$$(21) \quad \bar{X} = \frac{3.8 + 4.0 + 4.0}{3} = 3.933$$

$$(22) \quad s^2 = \frac{\sum (X - \bar{X})^2}{n-1} \\ = \frac{(3.8 - 3.933)^2 + (4.0 - 3.933)^2 + (4.0 - 3.933)^2}{3-1} = 0.013$$

The sample mean we calculated above estimates a population mean; and the sample variance estimates a population variance. *But that population is **not** a population of raw scores.*

It is now time to remind ourselves that these 3 scores are not raw scores; rather, they are the means of 3 equal-sized samples of raw scores. If the null hypothesis is true, it is convenient to think of the 3 samples as being drawn at random from the **same** population (rather than from 3 populations with the same mean and equal variances). In this case, the population distribution for which we have estimated the mean and variance is the **sampling distribution of the mean**:

$$\bar{X} \text{ calculated above} = \hat{\mu}_{\bar{X}} = \text{estimated mean of the Sampling Distribution of } \bar{X} \\ s^2 \text{ calculated above} = \hat{\sigma}_{\bar{X}}^2 = \text{estimated variance of the Sampling Distribution of } \bar{X}$$

The **central limit theorem** tells us (among other things) that the standard deviation of the sampling distribution of the mean, or the **standard error of the mean**, is equal to the population standard deviation divided by the square root of the sample size:

$$(23) \quad \hat{\sigma}_{\bar{X}} = s_x / \sqrt{n}$$

Taking the above, and squaring each term, we get:

$$(24) \quad \hat{\sigma}_{\bar{X}}^2 = s_x^2 / n \quad \text{then rearranging the terms,} \quad s_x^2 = n \hat{\sigma}_{\bar{X}}^2$$

Now, changing to more conventional notation for ANOVA, we get:

$$(25) \quad s_{\text{Between-groups}}^2 = n \hat{\sigma}_{\bar{X}}^2 = n \left(\frac{\sum_{j=1}^k (\bar{X}_j - \bar{X}_{\cdot})^2}{k-1} \right) \quad \{ \text{when all sample sizes are equal} \}$$

The following more general formula, in which the squared deviation of each treatment mean from the grand mean is multiplied by the size of that sample, should be used when sample sizes are not equal:

$$(26) \quad s_{\text{Between-groups}}^2 = \left(\frac{\sum_{j=1}^k n_j (\bar{X}_j - \bar{X}_{\cdot})^2}{k-1} \right)$$

Returning to the case of equal sample sizes [Equation (25)], **when the null hypothesis is true**, the variance of the sample means multiplied by the common sample size provides an estimate of the common population variance. This estimate is the **between-groups** variance estimate.

But if the null hypothesis were **false**, it would not be appropriate to describe the k samples as being drawn at random from the same population. In this case, the between-groups variance estimate would be an estimate of the population variance **plus** the effect of the independent variable, or the **treatment effect**. (That treatment effect is whatever variation there is among the population means.)

Putting it another way, the between-groups variance estimate always estimates the **population variance plus the treatment effect**. But when the null hypothesis is true, there is no real treatment effect (i.e., no variation among the population means), and so the between-groups variance estimate is simply an estimate of the common population variance.

$$(27) \quad s_{\text{Between-groups}}^2 = \hat{\sigma}_X^2 + \text{treatment effect} = \hat{\sigma}_X^2 + 0 \text{ when } H_0 \text{ is True}$$

Finally, if the reason for multiplying by n in equations (25) and (26) is still not clear, go back and take another look at Tables 3 and 5. Notice that in the $(\bar{Y}_j - \bar{Y}_\bullet)^2$ column, exactly the same squared deviation appears n times within each group. So to get the column sum, we could have summed only the **first** squared deviation in each group, and then multiplied by 5, the common sample size. This is what equation (25) does.

7.2 The within-groups variance estimate

As indicated earlier, homogeneity of variance is assumed, regardless of the truth or falsity of H_0 . The within-groups variance estimate estimates this common population variance, regardless of the truth or falsity of H_0 . It is calculated by pooling the variance estimates from the individual samples. When all samples are the same size, this amounts to simply taking the mean of the individual sample variances:

$$(28) \quad s_{\text{Within-groups}}^2 = s_{\text{Pooled}}^2 = \frac{\sum_{j=1}^k s_j^2}{k} = \hat{\sigma}_X^2$$

Note that equation (28) gives equal weight to each of the sample variances, and so is appropriate **only when all sample sizes are equal**. Here is a more general formula, which must be used when sample sizes are unequal, and can be used in any case:

$$(29) \quad s_{\text{Within-groups}}^2 = s_{\text{Pooled}}^2 = \frac{\sum_{j=1}^k SS_j}{\sum_{j=1}^k df_j} = \frac{SS_1 + SS_2 + \dots + SS_k}{df_1 + df_2 + \dots + df_k} = \frac{SS_{\text{Within-groups}}}{df_{\text{Within-groups}}} = \hat{\sigma}_X^2$$

Once we have these two variance estimates, everything else proceeds exactly as we saw in the first conceptual approach.

Plugging our data into equations (26), (29), and (17) yields the following:

$$(30) \quad s_{\text{Between-groups}}^2 = \left(\frac{\sum_{i=1}^k n_i (\bar{X}_i - \bar{X}_{\cdot})^2}{k-1} \right)$$

$$= \frac{5(3.8 - 3.933)^2 + 5(4.0 - 3.933)^2 + 5(4.0 - 3.933)^2}{3-1} = 0.067$$

$$(31) \quad s_{\text{Within}}^2 = \frac{SS_1 + SS_2 + SS_3}{df_1 + df_2 + df_3} = \frac{12.8 + 6 + 10}{4 + 4 + 4} = \frac{28.8}{12} = 2.400$$

$$(32) \quad F_{(2,12)} = \frac{0.067}{2.400} = 0.028$$

8. Assumptions of one-way ANOVA

Some of the assumptions of one-way ANOVA have already been mentioned in passing. Here is a more or less complete list:

- The dependent variable is measured on either an interval or ratio scale.
- Scores are sampled randomly from k normally distributed populations⁷.
- Homogeneity of variance: All population variances are assumed to be equal.
- All scores, and thus all errors in prediction, are independent of each other.

It should be noted that one-way ANOVA is quite robust⁸ to modest violations of the assumptions of normality and homogeneity of variance. Howell (1997, p.321) summarizes this topic nicely as follows:

⁷ Recall that each sample mean can be thought of as the predicted score for all subjects in that sample. Therefore, assuming normally distributed populations is equivalent to assuming that the errors in prediction (in the populations) are normally distributed. You may recall that this is an assumption of linear regression.

⁸ One definition of statistical robustness is "the resilience of your results, in the face of violations of your underlying assumptions." (I found this definition in a USENET newsgroup post, and do not know who first said it.)

In general, if the populations can be assumed to be symmetric, or at least similar in shape (e.g., all negatively skewed), and if the largest [sample] variance is no more than four times the smallest, the analysis of variance is most likely to be valid. It is important to note, however, that heterogeneity of variance and unequal sample sizes do not mix. If you have reason to anticipate unequal variances, make every effort to keep your sample sizes as equal as possible.

9. Review of the logic of one-way ANOVA

If the null hypothesis is true, the numerator and denominator of the F -ratio are simply two independent estimates of the common population variance, and so the ratio is expected to be roughly equal to 1. In this case, F will follow a theoretical F -distribution with $k-1$ and $N-k$ degrees of freedom (where k = the number of samples, and N = the total number of observations).

But if the null hypothesis is false, there will be a non-zero treatment effect in the numerator, and so F is expected to be greater than 1. If it is greater than 1 by a sufficient amount, we can reject the null hypothesis, and conclude that there is indeed a treatment effect—i.e., that the population means are not all equal. As you might expect, “greater than 1 by a sufficient amount” typically means that F is equal to or greater than the 95th percentile⁹ of the theoretical F -distribution with $k-1$ and $N-k$ degrees of freedom¹⁰. If you compute an F -ratio that is less than one, you know immediately that you cannot reject the null hypothesis of equal population means.

Variance Estimate	Condition of H_0	What it estimates
Between-groups	True	Common population variance + 0
	False	Common population variance + treatment effect
Within-groups	True or False	Common population variance

⁹ This is the case if you define statistical significance as $p \leq .05$. For $p \leq .01$, F would have to equal or exceed the 99th percentile of the theoretical F -distribution.

¹⁰ See pages 19 and 717 in Kleinbaum et al (1998) for more information on the theoretical F -distribution.

10. Variations in terminology

One of the challenges faced by students of statistics is that there are (at least) umpteen different sets of expressing the same basic idea. In the case of ANOVA, most of these variations have to do with the names used for the two estimates of variance. I referred to them as the between-groups and within-groups variance estimates. But there are a few other commonly used terms that you might encounter if you read another textbook, or look at websites on ANOVA.

I mentioned earlier in this chapter that *treatment* is often used in place of *between-groups* (e.g., $MS_{Treatment}$ instead of $MS_{Between}$), so no further comment is needed on that.

However, I should say a few words about why *error* and *residual* are often used instead of *within-groups* (e.g., MS_{error} or $MS_{residual}$ instead of MS_{within}). One way to conceive of one-way ANOVA is to say that we are trying to come up with a predicted score for every subject. As it turns out, the predicted score for each person is \bar{X}_j , the mean of the group to which that person belongs. The deviation of that person's actual score (or raw score) from their group mean is considered an error in prediction. So $\sum (Y_i - \bar{Y}_j)^2$ = the sum of the squared errors in prediction, = SS_{error} .

The explanation for *residual* is similar. When we look at a set of scores such as those shown in Tables 3 and 5, we see that not everyone has the same score. This is another way of saying that there is some variance. We then attempt to explain that variance—i.e., why doesn't everyone have the same score? One reason may be that not everyone received the same treatment. The variation of group means around the grand mean accounts for some portion of the total variation in scores. But unless there is no variation among people within groups, the variation among group means will not account for all of the variation in the scores. The portion of variance that is *left over* is the variation among people within groups—the within-groups variance. *Residual* is just another word for *left over*.

11. One-way ANOVA using SPSS

Nowadays, most analyses of variance are carried out using a statistical software package, such as SAS, Stata, or SPSS. The first thing you have to understand is how to structure your data file. For one-way ANOVA, most statistics packages require a data file with 2 columns, 1 to code group membership, and 1 to record the dependent variable (e.g., the columns labeled “Group” and “Y” in Table 5).

The following SPSS syntax (and associated output) shows three different ways to perform a one-way ANOVA on the data in Table 5. A similar syntax file (which uses a different data set) is available to those who want it at:

<http://www.angelfire.com/wv/bwhomedir/spss/anova1.SPS>

```

* -----
File:          anova1.sps
Author:        Bruce Weaver
Date:          December 30, 2000
Notes:         Second example in ANOVA1 chapter for HRM-723.
* -----

```

```

* NOTE:  The following was written with reference to SPSS version 9; it may
         be that some things are slightly different if you have an earlier
         or later version of the program .

```

```
DATA LIST LIST / numdist (f2.0) time (f5.0).
```

```
BEGIN DATA.
```

```

1      5
1      6
1      2
1      5
1      6
2      9
2      5
2      3
2      7
2      3
3      7
3     10
3     11
3     12
3     12

```

```
END DATA.
```

```
var lab
```

```

    numdist 'Number of distractors'
    time    'Time to solve problems'.

```

```
val lab numdist
```

```

1      '1 distractor'
2      '2 distractors'
3      '3 distractors'.

```

```
* List cases to show the structure of the data file .
```

```
list all.
```

```
List
```

```
NUMDIST  TIME
```

```

1      5
1      6
1      2
1      5
1      6
2      9
2      5
2      3
2      7
2      3
3      7
3     10
3     11
3     12
3     12

```

```
Number of cases read: 15    Number of cases listed: 15
```

```
* Note that there are 2 columns in the data file, 1 to code
```

group membership (must be a numeric code in SPSS),
and 1 to record the DV for that subject .

* There are several ways to perform a one-way ANOVA in SPSS.

* ----- Using MEANS Procedure to perform one-way ANOVA ----- .

* Perhaps the simplest is to use the /statistics=ANOVA subcommand
of the MEANS procedure, as follows.

means time by numdist /cells = count mean stddev variance /stat=anova.

Means

Case Processing Summary

	Cases					
	Included		Excluded		Total	
	N	Percent	N	Percent	N	Percent
Time to solve problems	15	100.0%	0	.0%	15	100.0%
* Number of distractors						

Report

Time to solve problems

Number of distractors	N	Mean	Std.	
			Deviation	Variance
1 distractor	5	4.80	1.64	2.700
2 distractors	5	5.40	2.61	6.800
3 distractors	5	10.40	2.07	4.300
Total	15	6.87	3.27	10.695

ANOVA Table

		Sum of Squares	df	Mean Square	F	Sig.
Time to solve problems	Between Groups (Combined)	94.533	2	47.267	10.275	.003
* Number of distractors	Within Groups	55.200	12	4.600		
Total		149.733	14			

Measures of Association

	Eta	Eta Squared
Time to solve problems	.795	.631
* Number of distractors		

* The eta-squared you see in the output is computed as SS_Between divided by SS_Total; it is sometimes called the "correlation ratio", and is similar to r-squared for a regression analysis; in other words, it indicates the proportion of total variance that is explained by the independent variable.

* ----- Using One-way ANOVA Procedure to perform one-way ANOVA ----- .

* If you look under Analyze-->Compare Means in the pull-down menus, you will see that the last option is One-way ANOVA; here is some syntax for that procedure (including descriptive statistics and a plot of the treatment means) .

```
ONEWAY
time BY numdist
/STATISTICS DESCRIPTIVES
/PLOT MEANS
/MISSING ANALYSIS .
```

Oneway

Descriptives

Time to solve problems

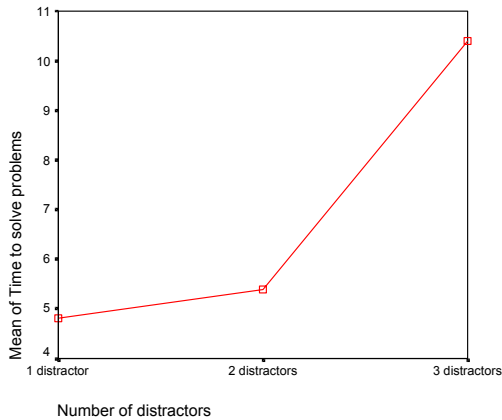
	N	Mean	Std. Deviation	Std. Error	95% Confidence Interval for Mean		Minimum	Maximum
					Lower Bound	Upper Bound		
1 distractor	5	4.80	1.64	.73	2.76	6.84	2	6
2 distractors	5	5.40	2.61	1.17	2.16	8.64	3	9
3 distractors	5	10.40	2.07	.93	7.83	12.97	7	12
Total	15	6.87	3.27	.84	5.06	8.68	2	12

ANOVA

Time to solve problems

	Sum of Squares	df	Mean Square	F	Sig.
Between Groups	94.533	2	47.267	10.275	.003
Within Groups	55.200	12	4.600		
Total	149.733	14			

Means Plots



* Eta-squared is not available in the output using this method; but a number of multiple comparison methods are available; some of these will be described in the next chapter of notes .

* ----- Using GLM UNIVARIATE Procedure to perform one-way ANOVA ----- .
 * The first option under Analyze-->General Linear Model in the pull-down menu is UNIVARIATE; here's how to perform a one-way ANOVA using GLM UNIVARIATE (again with descriptive statistics and a plot of means).

```
UNIANOVA
time BY numdist
/METHOD = SSTYPE(3)
/INTERCEPT = INCLUDE
/PLOT = PROFILE( numdist )
/EMMEANS = TABLES(numdist)
/PRINT = DESCRIPTIVE
/CRITERIA = ALPHA(.05)
/DESIGN = numdist .
```

Univariate Analysis of Variance

Between-Subjects Factors

	Value	Label	N
Number of distractors	1	1 distractor	5
	2	2 distractors	5
	3	3 distractors	5

Descriptive Statistics

Dependent Variable: Time to solve problems

Number of distractors	Mean	Std. Deviation	N
1 distractor	4.80	1.64	5
2 distractors	5.40	2.61	5
3 distractors	10.40	2.07	5
Total	6.87	3.27	15

Tests of Between-Subjects Effects

Dependent Variable: Time to solve problems

Source	Type III Sum of Squares	df	Mean Square	F	Sig.
Corrected Model	94.533 ^a	2	47.267	10.275	.003
Intercept	707.267	1	707.267	153.754	.000
NUMDIST	94.533	2	47.267	10.275	.003
Error	55.200	12	4.600		
Total	857.000	15			
Corrected Total	149.733	14			

a. R Squared = .631 (Adjusted R Squared = .570)

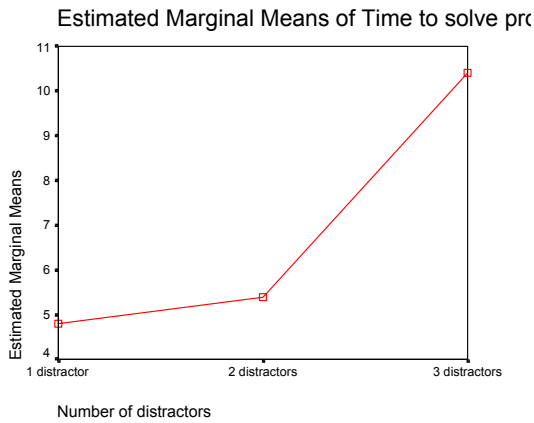
Estimated Marginal Means

Number of distractors

Dependent Variable: Time to solve problems

Number of distractors	Mean	Std. Error	95% Confidence Interval	
			Lower Bound	Upper Bound
1 distractor	4.800	.959	2.710	6.890
2 distractors	5.400	.959	3.310	7.490
3 distractors	10.400	.959	8.310	12.490

Profile Plots



- * Note that the ANOVA summary table produced by this procedure is a bit different--it has some extra rows we have not seen before.
- * The rows you have seen before are the ones labeled NUMDIST, ERROR and CORRECTED TOTAL; the SS, MS, and F values on these rows match what you saw on the BETWEEN, WITHIN, and TOTAL rows in the previous analyses .
- * You will also see an R-squared value of .631 below the ANOVA summary table; this is identical to the eta-squared value we saw earlier (in the output from the MEANS procedure); it indicates that 63.1% of the total variation in the DV is explained by the IV .
- * Finally, note that in all of the preceding output, the good folks at SPSS have use the column heading "Sig." where they should have use "p" .

* ===== .



References

- Howell, D.C. (2007). *Statistical methods for Psychology* (6th Ed.). Thomson-Wadsworth.
- Howell, D.C. (1997). *Statistical methods for Psychology* (4th Ed.). Belmont, CA: Duxbury.
- Kleinbaum, D.G., Kupper, L.L., Muler, K.E., & Nizam, A. (1998). *Applied regression analysis and other multivariable methods* (3rd Ed.). Pacific Grove, CA: Duxbury.

Acknowledgements

Thanks to Dave Evans for providing useful comments on a draft of this chapter.
