

## One-Way ANOVA

### 1. Purpose

Analysis of variance (ANOVA) is used when one wishes to determine whether two or more groups (e.g., classes A, B, and C) differ on some outcome of interest (e.g., an achievement test). The groups represent the IV and the outcome is the DV. The IV must be categorical (nominal) and the DV must be continuous (typically ordinal, interval, or ratio).

Some examples of when ANOVA would be appropriate include:

(a) Does a difference exist between males and females in terms of salary?

IV = sex (male, female); DV = salary (continuous variable)

(b) Do SAT scores differ among four different high schools?

IV = schools (school a, b, c, and d); DV = SAT scores

(c) Which teaching strategy is most effective in terms of mathematics achievement: jigsaw, peer tutoring, or writing to learn?

IV = teaching strategy (jigsaw, peer, write); DV = achievement scores

### 2. Hypotheses

Following with example (c), suppose one wanted to learn, via an experiment, which treatment was most effective, if any. The null hypothesis to be tested states that the groups (or treatment) are equivalent--no differences exists within the population among the group means. In symbolic form, the null is stated as:

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

where the mean achievement score for the jigsaw group is equal to  $\mu_1$ , the mean score for peer tutoring is  $\mu_2$ , and for writing to learn  $\mu_3$ .

The alternative hypothesis states that not all group means are equal; that is, one expects for the groups to have a different level of achievement. In symbolic form, the alternative is written as:

$$H_1: \mu_i \neq \mu_j \text{ for some } i, j;$$

or

$$H_1: \text{not all } \mu_j \text{ are equal.}$$

For hypotheses, the subscripts  $i$  and  $j$  simply represent different groups, like jigsaw ( $i$ ) and peer tutoring ( $j$ ). The alternative hypothesis *does not* specify which groups differ; rather, it simply indicates that at least two of the group means differ.

### 3. Why not separate t-tests?

Occasionally researchers will attempt to analyze the data from three (or more) groups via separate t-tests. That is, separate comparisons will be made as follows:

t-test #1:   jigsaw               vs.   write-to-learn  
 t-test #2:   jigsaw               vs.   peer-tutoring  
 t-test #3:   peer-tutoring       vs.   write-to-learn

The problem with such an analysis procedure is that the more separate t-tests one does, the more likely one is to commit a Type 1 error in hypothesis testing. Recall that a Type 1 error is rejecting the null hypothesis (i.e., no difference) when in fact it should not be rejected. In short, the more t-tests one performs, the more likely one will claim a difference exists when in fact it does not exist.

If one sets the probability of committing a Type 1 error ( $\alpha$ ) at .05 for each t-test performed, then with three separate t-tests of the same data, the probability of committing a Type 1 for one of the three tests is:

$$1 - (1 - \alpha)^C$$

where  $\alpha$  is the Type 1 error rate (usually set at .05) and  $C$  is the number of comparisons being made. For the three t-tests, the increased Type 1 error rate is:

$$\begin{aligned} &= 1 - (1 - .05)^3 \\ &= 1 - (0.95)^3 \\ &= 1 - .857375 \\ &= .142625 \end{aligned}$$

This increased likelihood of committing a Type 1 error is referred to as the *experimentwise error rate* or *familywise error rate*.

### 4. Linear Model Representation of ANOVA

The ANOVA model can be represented in the form of a linear model. For example, an individual student's score on the mathematical achievement test is symbolized as follows:

$$Y_{ij} = \alpha + \beta_j + e_{ij}$$

where

- $Y_{ij}$  = the  $i^{\text{th}}$  subject's score in the  $j^{\text{th}}$  group;
- $\alpha$  = the grand mean;
- $\beta_j = \beta_j - \alpha$  = the effect of belonging to group  $j$  (for those in EDUR 8132,  $\beta_j$  would be dummy variables); whether group means vary or differ from the grand mean
- $e_{ij}$  = 'random error' associated with this score.

The random errors, the  $e_{ij}$ , are expected (assumed) to be normally distributed with a mean of zero for each of the groups. The variance of these errors across the groups is  $\sigma_e^2$ , which is the population error variance. It is also expected (assumed) that the error variances for each group will be equal.

In terms of the linear model, ANOVA tests whether variability in  $\beta_j$  exists—that is, whether variability greater than one would expect *due to sampling error* among the group means exists.

### 5. Logic of Testing $H_0$ in ANOVA

The question of interest is whether the achievement means for the three groups differ, statistically, from one-another. If the means differ, then there will be variation among (*between* in ANOVA parlance) the means, so

one could calculate the amount of variation between the means in terms of a variance. The variance between the means is symbolized as  $s_j^2$ .

In addition to the variance between the group means, there will also be variance among individual scores within each group. That is, each student studying mathematics via the jigsaw method will likely score slightly differently from a peer who also studied mathematics via the jigsaw method. Thus, there will also be some variance *within* each group due to differences among individual scores.

The schematic below illustrates variation *between* group means and variation *within* groups.

Table 1: Between and Within Sources of Variation Sample Data

Source of Variation		Group 3: Write to Learn	Group 1: Jigsaw	Group 2: Peer Tutor
<u>Variation Between</u>	(Group Means, $\bar{X}_j$ )	75	85	95
<u>Variation Within</u>	(Individual Scores)	77 73 78 72	88 84 82 86	95 96 94 95

Note. Grand Mean,  $\bar{X} = \bar{X}_{..} = 85$ .

In short, if the amount of variation *between* groups is larger than the amount of variation *within* groups, then the groups are said to be statistically different.

Thus, one is considering the ratio of the between groups variation to the within groups variation -- between/within.

### 5a. ANOVA Computation

ANOVA computation is based upon the information found in the summary table below.

One-way ANOVA Summary Table				
Source	SS	df	MS	F
between	$SS_b$	$df_b (df_1) = j - 1$	$SS_b/df_b$	$MS_b/MS_w$
within	$SS_w$	$df_w (df_2) = n - j$	$SS_w/df_w$	
total	$SS_t$	$df_t = n - 1$		

MS (mean squares) is just another term for Variance, variance between and variance within groups.

F is the ratio of between variance to within variance. The greater the group separation, the greater will be the F ratio, the more the groups overlap and are therefore indistinguishable, the small the F ratio.

Here is an illustration to demonstrate group separation:

<http://tinyurl.com/24kpg5j>

or here

<https://docs.google.com/drawings/edit?id=17eS69paOqp3G6Ejl8L4wj1YjX-NS50jiqjx20z6mUc&hl=en&authkey=CLne2ZIB>

Other illustrations, with included calculated F ratio, are here:

<http://www.psychology.heacademy.ac.uk/miniprojects/anova/anova1.html> (link dead 6/12)

<http://www.buseco.monash.edu.au/mkt/resources/applets/one-way-anova.html>

[http://bcs.whfreeman.com/ips4e/cat\\_010/applets/anova.html](http://bcs.whfreeman.com/ips4e/cat_010/applets/anova.html)

### 5b. Sums of Squares

The amount of variation between and within groups in ANOVA is determined by *sums of squares* (SS), and SS are used to test the null hypothesis of no group differences. SS can be partitioned into three components--total, within, and between--as illustrated below.

$$SS_{\text{total}} = SS_{\text{between}} + SS_{\text{within}}, \text{ or}$$

$$SS_t = SS_b + SS_w.$$

$SS_{\text{total}}$

The total SS is defined as follows:

$$SS_t = \sum_{j=1}^j \sum_{i=1}^{n_j} (X_{ij} - \bar{X})^2$$

where  $i$  represents each individual's score,  $j$  represents the unique groups, and  $\bar{X}$  is the grand mean across all individuals and groups. Also,  $\sum_{j=1}^j$  means that scores across each group,  $j$ , must be summed, and  $\sum_{i=1}^{n_j}$  indicates that individual scores within each group  $j$  must be summed, i.e., summing occurs within and across groups.

Using the data provided above in Table 1, the total sums of squares would be:

$$\begin{aligned} SS_t &= (77 - 85)^2 + (73 - 85)^2 + (78 - 85)^2 + (72 - 85)^2 + \\ &\quad (88 - 85)^2 + (84 - 85)^2 + (82 - 85)^2 + (86 - 85)^2 + \\ &\quad (95 - 85)^2 + (96 - 85)^2 + (94 - 85)^2 + (95 - 85)^2 \\ &= (-8)^2 + (-12)^2 + (-7)^2 + (-13)^2 + \\ &\quad (3)^2 + (-1)^2 + (-3)^2 + (1)^2 + \\ &\quad (10)^2 + (11)^2 + (9)^2 + (10)^2 \\ &= (64) + (144) + (49) + (169) + \\ &\quad (9) + (1) + (9) + (1) + \\ &\quad (100) + (121) + (81) + (100) \\ &= 848 \end{aligned}$$

Recall the formula for the sample variance:

$$s^2 = \frac{\Sigma(X - \bar{X})^2}{n-1}.$$

If  $SS_t$  is divided by  $n - 1$ , then the product will equal the variance for the data. For example,

$$SS_t = 848,$$

so the variance for this data is

$$s^2 = 848/11 = 77.09.$$

**$SS_{\text{between}}$  (also  $SS_{\text{regression}}$  or  $SS_{\text{treatment}}$ )**

The sums of squares *between* is defined as:

$$SS_b = \sum n_j (\bar{X}_j - \bar{X})^2$$

where  $\bar{X}_j$  is the mean for the  $j^{\text{th}}$  group,  $\bar{X}$  is the grand mean, and  $n_j$  is the sample size for each group.

To illustrate,  $SS_b$  for the sample data in Table 1 is:

$$\begin{aligned} SS_b &= 4(75 - 85)^2 + 4(85 - 85)^2 + 4(95 - 85)^2 \\ &= 4(-10)^2 + 4(0)^2 + 4(10)^2 \\ &= 4(100) + 4(0) + 4(100) \\ &= 800. \end{aligned}$$

**$SS_{\text{within}}$  (also  $SS_{\text{error}}$ ,  $SS_e$ , or  $SS_{\text{residual}}$ )**

The sums of squares within is defined as:

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

where  $\bar{X}_j$  is the mean for the  $j^{\text{th}}$  group, and  $X_{ij}$  is the raw, individual score for the  $i^{\text{th}}$  person in the  $j^{\text{th}}$  group. Note the subtle difference between the formula for  $SS_t$  and  $SS_w$ ; in  $SS_t$  raw scores are subtracted from the overall mean,  $\bar{X}$ , while in  $SS_w$  raw scores are subtracted from each group's mean,  $\bar{X}_j$ .

To illustrate,  $SS_w$  for the sample data is:

$$\begin{aligned} SS_w &= (77 - 75)^2 + (73 - 75)^2 + (78 - 75)^2 + (72 - 75)^2 + \\ &\quad (88 - 85)^2 + (84 - 85)^2 + (82 - 85)^2 + (86 - 85)^2 + \\ &\quad (95 - 95)^2 + (96 - 95)^2 + (94 - 95)^2 + (95 - 95)^2 \\ &= (2)^2 + (-2)^2 + (3)^2 + (-3)^2 + \\ &\quad (3)^2 + (-1)^2 + (-3)^2 + (1)^2 + \\ &\quad (0)^2 + (1)^2 + (-1)^2 + (0)^2 \\ &= (4) + (4) + (9) + (9) + \end{aligned}$$

$$(9) + (1) + (9) + (1) + \\ (0) + (1) + (1) + (0)$$

$$= 48$$

The total SS,  $SS_t$ , equals the combination of  $SS_b$  and  $SS_w$ . For the data given above,  $SS_t = 848$ ,  $SS_b = 800$ , and  $SS_w = 48$ ; thus,

$$SS_t = SS_b + SS_w.$$

$$848 = 800 + 48$$

$$848 = 848.$$

### 5c. Degrees of Freedom (df)

ANOVA, like the t-test, has what is called df. DF simply indicate the amount of information in the data that can be used to estimate parameters once certain restrictions to the data are applied. It is not important that you understand this concept thoroughly—what is important is that you understand how to apply df in ANOVA.

In ANOVA there are two sets of df: df between and df within. Each are described below.

#### **df<sub>between</sub>, df<sub>b</sub>, df<sub>1</sub>, or v<sub>b</sub>**

The  $df_b$  are calculated as:

$$df_b (\underline{v}_b) = j - 1$$

where j is the number of groups in the study. In the current example, the between df are

$$df_b = j - 1 = 3 - 1 = 2.$$

#### **df<sub>within</sub>, df<sub>w</sub>, df<sub>2</sub>, or v<sub>w</sub>**

The  $df_w$  are calculated as:

$$df_w (\underline{v}_w) = n - j$$

where n is the total number of subjects in the sample. In the current example, within df are

$$df_w = n - j = 12 - 3 = 9.$$

### 5d. Estimating Variances Between and Within

When the various SS (i.e.,  $SS_b$  and  $SS_w$ ) are divided by their appropriate dfs (i.e.,  $df_b$  and  $df_w$ ), the result is an estimate of the variance between and within. These variances are called mean squares in ANOVA.

#### **Mean Square Between, MS<sub>b</sub> or MS<sub>regression</sub>**

The mean square between, an estimate of the between variance, is calculated as

$$MS_b = SS_b / df_b.$$

In the current example the  $MS_b$  is

$$MS_b = 800 / 2 = 400.$$

#### **Mean Square Within, MS<sub>w</sub> or MS<sub>error</sub>**

The mean square within, an estimate of the within variance, is calculated as

$$MS_w = SS_w / df_w.$$

In the current example the  $MS_w$  is

$$MS_w = 48 / 9 = 400/5.33 = 75.05.$$

### 5e. Testing $H_0$ in ANOVA: The F test

Determining whether the groups have significantly different means requires the comparison of the between variation,  $MS_b$ , to the within variation,  $MS_w$ . As previously noted, if the between variation is substantially larger than the within variation, then the null hypothesis of no difference is rejected in favor of the alternative hypothesis that at least one of the group means is different from the others.

The ratio  $MS_b/MS_w$  is labeled the F ratio. The F ratio ranges from 0 to positive infinity. The larger the F ratio (provided F is larger than 1), the more likely the group means are different and that this difference cannot be attributed to *sampling error*.

#### Assessing the F ratio for Statistical Significance

The F ratio obtained from the one-way ANOVA has degrees of freedom like that of the t-test. The df for the F ratio are represented by two components,  $df_b$  and  $df_w$ .

For the current example, the F ratio is

$$F = MS_b/MS_w = 400/5.333 = 75.$$

This calculated F ratio of 75 is compared against a critical F value that can be located in the F table (see Table 2, below).

The critical F values are found by locating the point at which the column df,  $df_b$ , intersects the row df,  $df_w$ . The column represents the  $df_b$ , and the row represents the  $df_w$ .

In the example data, the critical F value is ( $\alpha = .05$ )

$$F_{df_1, df_2} = F_{2,9} = 4.26.$$

In order to determine whether the null hypothesis of no difference should be rejected, the calculated F ratio is compared to the critical F ratio. The following decision rule assists in determining whether the null should be rejected:

**If  $F \geq F_{crit}$ , then reject  $H_0$ , otherwise fail to reject  $H_0$ ;**

where F is the calculated F value obtained from the data, and  $F_{crit}$  is the critical F valued taken from the table.

For the example data, the decision rule is:

**If  $75 \geq 4.26$ , then reject  $H_0$ , otherwise fail to reject  $H_0$ ;**

Since 75 is greater than 4.26,  $H_0$  is rejected and one concludes that the group means are not equal in the population.

Tabled critical F values can be found here:

[http://www.bwgriffin.com/gsu/courses/edur8132/notes/critical\\_f\\_values.pdf](http://www.bwgriffin.com/gsu/courses/edur8132/notes/critical_f_values.pdf)

A second method of testing  $H_0$  is via p-values, and this will be the method used most frequently. Each and every test statistic (e.g., t, F, r) has associated with it a probability value, or p-value. For the F test, the p-value represents *the probability of observing mean differences as large as that observed, or larger, in the given data assuming that  $H_0$  is true*—i.e., that no mean differences exist in the population.

In hypothesis testing, one always assumes the null hypothesis is true. Given this, then how likely is it that mean differences like those found in the sample data will occur given a random sample. The larger the p-value (e.g.,  $p = .50$ ,  $.33$ , or  $.19$ ), the more likely one can conclude that the mean differences observed can be attributed to sampling error rather than to true or real group differences.

If, however, the p-value is small (e.g.,  $.05$ ,  $.01$ , or  $.02$ ), the less likely the group differences observed could have resulted from random sampling error, and the more likely it is that the differences are real. The decision rule for p-values is:

**If  $p\text{-value} \leq \alpha$ , then reject  $H_0$ , otherwise fail to reject  $H_0$ ;**

where p-value is the probability value obtained from the computer, and  $\alpha$  is the alpha level the researcher sets (such as  $.05$  or  $.01$ ).

The p-value obtained for the sample data was  $.000$ . (Actually, a p-value cannot equal  $.000$ , but in some software precision is limited so p-values are reported as zero.) The decision rule is:

**If  $.000 \leq .05$ , then reject  $H_0$ , otherwise fail to reject  $H_0$ ;**

so  $H_0$  is rejected.



## 5f. Online Demonstrations of ANOVA

[http://onlinestatbook.com/stat\\_sim/one\\_way/index.html](http://onlinestatbook.com/stat_sim/one_way/index.html)

<http://www.buseco.monash.edu.au/mkt/resources/applets/one-way-anova.html>

## 6. SPSS ONEWAY Commander Output for Sample Data (To be illustrated in class for EDUR 8131)

SPSS is commonly used statistical software in research. The following are results obtained from SPSS for the sample data.

```

* * *  A N A L Y S I S   O F   V A R I A N C E   * * *

      DV
    by  IV

      UNIQUE sums of squares
      All effects entered simultaneously

Source of Variation           Sum of      Mean      F      Sig
                             Squares      Square      of F   of F
Main Effects
  IV                          800.000      400.000    75.000  .000
                             800.000      400.000    75.000  .000
Explained                    800.000      400.000    75.000  .000
Residual                      48.000       5.333
Total                        848.000     77.091

```

Another SPSS results section from SPSS follows:

```

- - - - -  O N E W A Y  - - - - -

      Variable  DV
    By Variable  IV

                        Analysis of Variance

Source                D.F.      Sum of      Mean      F      F
Between Groups        2         800.0000   400.0000   75.0000  .0000
Within Groups         9         48.0000    5.3333
Total                 11        848.0000

Group   Count   Mean   Standard  Standard
                Mean   Deviation  Error
Grp 1    4    75.0000  2.9439    1.4720
Grp 2    4    85.0000  2.5820    1.2910
Grp 3    4    95.0000  .8165     .4082
Total   12    85.0000  8.7801    2.5346
          79.4214  TO
          90.5786

GROUP   MINIMUM   MAXIMUM
Grp 1   72.0000    78.0000
Grp 2   82.0000    88.0000
Grp 3   94.0000    96.0000
TOTAL   72.0000    96.0000

```

## 7. Multiple Comparisons

If more than two groups are compared, and if the F-ratio for the IV is statistically significant, one often will perform pairwise comparisons (each group's mean compared against another) or more complex contrasts (e.g., the mean of groups A and B vs. group C).

For those in EDUR 8132 this has already been covered in regression and will be illustrated via SPSS during chats.

For those in EDUR 8131 this will be covered separately. Presentation below assumes multiple comparisons with either the Bonferroni or Scheffé correction procedure discussed.

## 8. SPSS Out Using GENERAL LINEAR MODEL->Univariate Command

Steps in SPSS:

1. Enter data with IV defined as Group and DV as Scores. Group may be numeric or string variable. Below is copy of data as entered in SPSS.

Group	Scores
write	77.00
write	73.00
write	78.00
write	72.00
jigsaw	88.00
jigsaw	84.00
jigsaw	82.00
jigsaw	86.00
peertutor	95.00
peertutor	96.00
peertutor	94.00
peertutor	95.00

2. Select following commands in SPSS:

- a. Analyze->General Linear Model->Univariate
- b. Move Group to "Fixed Factors" box, move Scores to "Dependent Variables" box
- c. Select "Options" and in new window select "Group" from "Factors and Factor Interactions" window and move to "Display Means for" box; next place mark next to "Compare Main Effects" and select "Bonferroni" in pull-down menu; next place mark next to "Descriptive Statistics"; next click "Continue"
- d. Click "Ok"

Results are displayed below. These will be discussed during chats.

**Descriptive Statistics**

Dependent Variable: scores

group	Mean	Std. Deviation	N
jigsaw	85.0000	2.58199	4
peertutor	95.0000	.81650	4
write	75.0000	2.94392	4
Total	85.0000	8.78014	12

**Tests of Between-Subjects Effects**

Dependent Variable: scores

Source	Type III Sum of Squares	df	Mean Square	F	Sig.
Corrected Model	800.000(a)	2	400.000	75.000	.000
Intercept	86700.000	1	86700.000	16256.250	.000
group	800.000	2	400.000	75.000	.000
Error	48.000	9	5.333		
Total	87548.000	12			
Corrected Total	848.000	11			

a R Squared = .943 (Adjusted R Squared = .931)

**Estimates**

Dependent Variable: scores

group	Mean	Std. Error	95% Confidence Interval	
			Lower Bound	Upper Bound
jigsaw	85.000	1.155	82.388	87.612
peertutor	95.000	1.155	92.388	97.612
write	75.000	1.155	72.388	77.612

**Pairwise Comparisons**

Dependent Variable: scores

(I) group	(J) group	Mean Difference (I-J)	Std. Error	Sig.(a)	95% Confidence Interval for Difference(a)	
					Lower Bound	Upper Bound
jigsaw	peertutor	-10.000(*)	1.633	.001	-14.790	-5.210
	write	10.000(*)	1.633	.001	5.210	14.790
peertutor	jigsaw	10.000(*)	1.633	.001	5.210	14.790
	write	20.000(*)	1.633	.000	15.210	24.790
write	jigsaw	-10.000(*)	1.633	.001	-14.790	-5.210
	peertutor	-20.000(*)	1.633	.000	-24.790	-15.210

Based on estimated marginal means

\* The mean difference is significant at the .05 level.

a Adjustment for multiple comparisons: Bonferroni.

## 8. APA Styled Results

Sometimes researchers report only the results of the F test in manuscripts, and such a report may resemble the following:

Results from the achievement test indicate that type of supplemental instruction does appear to influence achievement. The results show a statistically significant difference in achievement scores ( $F_{2,9} = 75, p < .05$ ) across the three groups.

Alternative methods for reporting the calculated F ratio include:

- (a)  $F = 75, df = 2/9, p < .05$
- (b)  $F_{(2,9)} = 75, p < .05$
- (c)  $F(2, 9) = 75, p = .000$ .

A better approach, and the one required in this course, is to provide more detail as illustrated below.

*Table 1*  
*ANOVA Results and Descriptive Statistics for Mathematics Scores by Type of Instruction*

School Type	Mean	SD	n
Write to Learn	75.00	2.94	4
Jigsaw	85.00	2.58	4
Peer Tutor	95.00	0.82	4

  

Source	SS	df	MS	F
Instruction	800.00	2	400.00	75.00*
Error	48.00	9	5.33	

Note.  $R^2 = .94$ , adj.  $R^2 = .93$ .

\*  $p < .05$

*Table 2*  
*Multiple Comparisons and Mean Differences in Mathematics Scores by Type of Instruction*

Comparison	Mean Difference	s.e.	95% CI
Write vs. Jigsaw	-10.00*	1.63	-14.79, -5.21
Write vs. Peer Tutor	-20.00*	1.63	-24.79, -15.21
Peer Tutor vs. Jigsaw	10.00*	1.63	5.21, 14.79

\*  $p < .05$ , where p-values are adjusted using the Bonferroni method.

Statistical analysis of mathematics scores show statistically significant mean differences, at the .05 level, among the three groups examined. Multiple comparisons results show that those in peer tutoring demonstrated the highest performance, those in jigsaw the second highest, and those in write to learn the lowest. Each pairwise comparison was statistically significant.

Scheffé confidence intervals also may be obtained for one-way ANOVA in SPSS, or one may opt to use the spreadsheet for CI linked on the course web site.

## 9. Assumptions of ANOVA

In order that calculated probabilities for the F test be as accurate as possible, certain assumptions are needed. The assumptions for the F test, and hence ANOVA, are:

1. The observations (scores) within each group are normally distributed about their group mean,  $\mu_j$ , with a variance of  $\sigma_w^2 = \sigma_\varepsilon^2 = \sigma^2$
2. Each group is expected to have the same variance,  $\sigma^2$ , and this assumption is referred to as *homogeneity of variance*.
3. The observations (scores) within each group are independently distributed about their group mean.

Often these assumptions are expressed with the following notation:

$$\varepsilon_{ij} \text{ (or } Y_{ij}) \sim \text{NID}(0, \sigma^2)$$

where NID means *normally and independently distributed* with mean of 0 and variance of  $\sigma_\varepsilon^2$ .

Three possible violations of these assumptions are possible. Each violation threatens the validity of inferences with regard to Type 1 and Type 2 errors in hypothesis testing using the F test. In short, with each violation, one is unable to determine precisely the correct p-value for a given calculated F or the correct statistical *power* level for the test. The three possible violations are (G. Glass & S. Hopkins, 1984, *Statistical methods in education and psychology*, p. 351):

- (1) non-normality
- (2) heterogeneity of variance among groups
- (3) non-independence.

Non-normality is typically not problematic due to the central limit theorem when group sizes (n's) are approximately equal. Also, when n's are approximately equal, heterogeneity of variances has little impact on p-values or power.

When the assumption of independence is violated, all statements regarding statistical inferences are incorrect. Only when dependence is the result of matched pairs (as in the correlated t-test) can violation of this assumption be taken into account. Typically lack of independence results from non-random sampling, such as convenience sampling in which intact groups, like classrooms, are used.

Glass and Hopkins (1984, see above citations) provide a much more detailed discussion of the impact of assumption violations upon statistical inference (pp. 352 - 357).

## 10. Exercises

1. Does a difference exist between men and women in terms of salary for assistant professors at GSU?

Men	Women
24500	39000
30500	37500
27000	29000
30500	32500
32500	34500

### Answer

SPSS output:

```

- - - - - O N E W A Y - - - - -

Variable  SALARY
By Variable  SEX

Analysis of Variance

Source          D.F.      Sum of Squares      Mean Squares      F Ratio      F Prob.
Between Groups      1      75625000.00      75625000.00      5.7895      .0428
Within Groups      8      104500000.0      13062500.00
Total              9      180125000.0

Group   Count      Mean      Standard Deviation      Standard Error      95 Pct Conf Int for Mean
Grp 0      5      34500.0000      3984.3444      1781.8530      29552.8634 TO 39447.1366
Grp 1      5      29000.0000      3201.5621      1431.7821      25024.8002 TO 32975.1998
Total     10      31750.0000      4473.6885      1414.7045      28549.7177 TO 34950.2823

GROUP      MINIMUM      MAXIMUM
Grp 0      29000.0000      39000.0000
Grp 1      24500.0000      32500.0000

TOTAL      24500.0000      39000.0000

```

It appears that assistant professors' salaries do differ, statistically, by sex ( $F[1, 8] = 5.79$ ,  $p = .042$ ) with females having the higher salary ( $M_{\text{female}} = \$34,500$ ;  $M_{\text{male}} = \$29,000$ ).

2. A researcher is interested in learning whether frequency of reading at home to elementary-aged children produces differential effects on reading achievement. After obtaining information from a randomly selected sample of parents about this behavior, the following classifications and standardized achievement scores were recorded. (Note: frequency classifications as follows: a = less than once per month, b = once to three times per month, c = more than three times per month.)

Freq. of Reading	Achievement
a	48
a	37
a	47
a	65
b	57
b	39
b	49
b	45
c	61
c	55
c	51
c	30

Is frequency of reading at home related to student reading achievement?

Answer

SPSS output:

```

* * * A N A L Y S I S   O F   V A R I A N C E * * *

      ACH
by    FREQ

      UNIQUE sums of squares
      All effects entered simultaneously
    
```

Source of Variation	Sum of Squares	DF	Mean Square	F	Sig of F
Main Effects	8.167	2	4.083	.033	.968
FREQ	8.167	2	4.083	.033	.968
Explained	8.167	2	4.083	.033	.968
Residual	1120.500	9	124.500		
Total	1128.667	11	102.606		

There is no statistical evidence that frequency of reading at home results in differential levels of reading achievement ( $F[2, 9] = 0.03, p = .97$ ).