Statistics Today

Is Seeing Really Believing?

Many adults look on the eyewitness testimony of children with skepticism. They believe that young witnesses' testimony is less accurate than the testimony of adults in court cases. Several statistical studies have been done on this subject.

In a preliminary study, three researchers selected fourteen 8-year-olds, fourteen 12-year-olds, and fourteen adults. The researchers showed each group the same video of a crime being committed. The next day, each witness responded to direct and cross-examination questioning. Then the researchers, using statistical methods explained in this chapter, were able to determine if there were differences in the accuracy of the testimony of the three groups on direct examination and on cross-examination. The statistical methods used here differ from the ones explained in Chapter 9 because there are three groups rather than two. See Statistics Today—Revisited at the end of this chapter.

Source: C. Luus, G. Wells, and J. Turtle, "Child Eyewitnesses: Seeing Is Believing," *Journal of Applied Psychology* 80, no. 2, pp. 317–26.

Historical Note

The methods of analysis of variance were developed by R. A. Fisher in the early 1920s.

Introduction

The F test, used to compare two variances as shown in Chapter 9, can also be used to compare three or more means. This technique is called *analysis of variance*, or *ANOVA*. It is used to test claims involving three or more means. (*Note:* The F test can also be used to test the equality of two means. But since it is equivalent to the t test in this case, the t test is usually used instead of the F test when there are only two means.) For example, suppose a researcher wishes to see whether the means of the time it takes three groups of students to solve a computer problem using Fortran, Basic, and Pascal are different. The researcher will use the ANOVA technique for this test. The z and t tests should not be used when three or more means are compared, for reasons given later in this chapter.

For three groups, the F test can only show whether a difference exists among the three means. It cannot reveal where the difference lies—that is, between $\bar{X}_1$ and $\bar{X}_2$, or $\bar{X}_1$ and $\bar{X}_3$, or $\bar{X}_2$ and $\bar{X}_3$. If the F test indicates that there is a difference among the means, other statistical tests are used to find where the difference exists. The most commonly used tests are the Scheffé test and the Tukey test, which are also explained in this chapter.

The analysis of variance that is used to compare three or more means is called a *one-way analysis of variance* since it contains only one variable. In the previous example, the variable is the type of computer language used. The analysis of variance can be extended to studies involving two variables, such as type of computer language used and mathematical background of the students. These studies involve a *two-way analysis of variance.* Section 12–3 explains the two-way analysis of variance.

12–1 One-Way Analysis of Variance

Objective 1

Use the one-way ANOVA technique to determine if there is a significant difference among three or more means.

When an F test is used to test a hypothesis concerning the means of three or more populations, the technique is called **analysis of variance** (commonly abbreviated as **ANOVA**). At first glance, you might think that to compare the means of three or more samples, you can use the t test, comparing two means at a time. But there are several reasons why the t test should not be done.

First, when you are comparing two means at a time, the rest of the means under study are ignored. With the F test, all the means are compared simultaneously. Second, when you are comparing two means at a time and making all pairwise comparisons, the probability of rejecting the null hypothesis when it is true is increased, since the more t tests that are conducted, the greater is the likelihood of getting significant differences by chance alone. Third, the more means there are to compare, the more t tests are needed. For example, for the comparison of 3 means two at a time, 3 t tests are required. For the comparison of 5 means two at a time, 10 tests are required. And for the comparison of 10 means two at a time, 45 tests are required.

Assumptions for the F Test for Comparing Three or More Means

1. The populations from which the samples were obtained must be normally or approximately normally distributed.
2. The samples must be independent of one another.
3. The variances of the populations must be equal.

Even though you are comparing three or more means in this use of the F test, *variances* are used in the test instead of means.

With the F test, two different estimates of the population variance are made. The first estimate is called the **between-group variance,** and it involves finding the variance of the means. The second estimate, the **within-group variance,** is made by computing the variance using all the data and is not affected by differences in the means. If there is no difference in the means, the between-group variance estimate will be approximately equal to the within-group variance estimate, and the F test value will be approximately equal to 1. The null hypothesis will not be rejected. However, when the means differ significantly, the between-group variance will be much larger than the within-group variance; the F test value will be significantly greater than 1; and the null hypothesis will be rejected. Since variances are compared, this procedure is called *analysis of variance* (ANOVA).

For a test of the difference among three or more means, the following hypotheses should be used:

H_0: $\mu_1 = \mu_2 = \cdots = \mu_k$

H_1: At least one mean is different from the others.

As stated previously, a significant test value means that there is a high probability that this difference in means is not due to chance, but it does not indicate where the difference lies.

The degrees of freedom for this F test are d.f.N. $= k - 1$, where k is the number of groups, and d.f.D. $= N - k$, where N is the sum of the sample sizes of the groups $N = n_1 + n_2 + \cdots + n_k$. The sample sizes need not be equal. The F test to compare means is always right-tailed.

Examples 12–1 and 12–2 illustrate the computational procedure for the ANOVA technique for comparing three or more means, and the steps are summarized in the Procedure Table shown after the examples.

Example 12–1

Lowering Blood Pressure

A researcher wishes to try three different techniques to lower the blood pressure of individuals diagnosed with high blood pressure. The subjects are randomly assigned to three groups; the first group takes medication, the second group exercises, and the third group follows a special diet. After four weeks, the reduction in each person's blood pressure is recorded. At $\alpha = 0.05$, test the claim that there is no difference among the means. The data are shown.

Medication	Exercise	Diet
10	6	5
12	8	9
9	3	12
15	0	8
13	2	4
$\bar{X}_1 = 11.8$	$\bar{X}_2 = 3.8$	$\bar{X}_3 = 7.6$
$s_1^2 = 5.7$	$s_2^2 = 10.2$	$s_3^2 = 10.3$

Solution

Step 1 State the hypotheses and identify the claim.

H_0: $\mu_1 = \mu_2 = \mu_3$ (claim)

H_1: At least one mean is different from the others.

Step 2 Find the critical value. Since $k = 3$ and $N = 15$,

d.f.N. $= k - 1 = 3 - 1 = 2$

d.f.D. $= N - k = 15 - 3 = 12$

The critical value is 3.89, obtained from Table H in Appendix C with $\alpha = 0.05$.

Step 3 Compute the test value, using the procedure outlined here.

a. Find the mean and variance of each sample (these values are shown below the data).

b. Find the grand mean. The *grand mean,* denoted by $\bar{X}_{GM}$, is the mean of all values in the samples.

$$\bar{X}_{GM} = \frac{\Sigma X}{N} = \frac{10 + 12 + 9 + \cdots + 4}{15} = \frac{116}{15} = 7.73$$

When samples are equal in size, find $\bar{X}_{GM}$ by summing the $\bar{X}$'s and dividing by k, where $k =$ the number of groups.

c. Find the between-group variance, denoted by s_B^2.

$$s_B^2 = \frac{\Sigma n_i(\overline{X}_i - \overline{X}_{GM})^2}{k-1}$$

$$= \frac{5(11.8 - 7.73)^2 + 5(3.8 - 7.73)^2 + 5(7.6 - 7.73)^2}{3-1}$$

$$= \frac{160.13}{2} = 80.07$$

Note: This formula finds the variance among the means by using the sample sizes as weights and considers the differences in the means.

d. Find the within-group variance, denoted by s_W^2.

$$s_W^2 = \frac{\Sigma(n_i - 1)s_i^2}{\Sigma(n_i - 1)}$$

$$= \frac{(5-1)(5.7) + (5-1)(10.2) + (5-1)(10.3)}{(5-1) + (5-1) + (5-1)}$$

$$= \frac{104.80}{12} = 8.73$$

Note: This formula finds an overall variance by calculating a weighted average of the individual variances. It does not involve using differences of the means.

e. Find the F test value.

$$F = \frac{s_B^2}{s_W^2} = \frac{80.07}{8.73} = 9.17$$

Step 4 Make the decision. The decision is to reject the null hypothesis, since $9.17 > 3.89$.

Step 5 Summarize the results. There is enough evidence to reject the claim and conclude that at least one mean is different from the others.

Interesting Facts

The weight of 1 cubic foot of wet snow is about 10 pounds while the weight of 1 cubic foot of dry snow is about 3 pounds.

The numerator of the fraction obtained in step 3, part *c*, of the computational procedure is called the **sum of squares between groups,** denoted by SS_B. The numerator of the fraction obtained in step 3, part *d*, of the computational procedure is called the **sum of squares within groups,** denoted by SS_W. This statistic is also called the *sum of squares for the error.* SS_B is divided by d.f.N. to obtain the between-group variance. SS_W is divided by $N - k$ to obtain the within-group or error variance. These two variances are sometimes called **mean squares,** denoted by MS_B and MS_W. These terms are used to summarize the analysis of variance and are placed in a summary table, as shown in Table 12–1.

Table 12–1 Analysis of Variance Summary Table

Source	Sum of squares	d.f.	Mean square	F
Between	SS_B	$k - 1$	MS_B	
Within (error)	SS_W	$N - k$	MS_W	
Total				

Unusual Stat

The *Journal of the American College of Nutrition* reports that a study found no correlation between body weight and the percentage of calories eaten after 5:00 P.M.

In the table,

SS_B = sum of squares between groups

SS_W = sum of squares within groups

k = number of groups

$N = n_1 + n_2 + \cdots + n_k$ = sum of sample sizes for groups

$$MS_B = \frac{SS_B}{k-1}$$

$$MS_W = \frac{SS_W}{N-k}$$

$$F = \frac{MS_B}{MS_W}$$

The totals are obtained by adding the corresponding columns. For Example 12–1, the **ANOVA summary table** is shown in Table 12–2.

12–2 Analysis of Variance Summary Table for Example 12-1

Source	Sum of squares	d.f.	Mean square	F
Between	160.13	2	80.07	9.17
Within (error)	104.80	12	8.73	
Total	264.93	14		

Most computer programs will print out an ANOVA summary table.

Example 12–2

Employees at Toll Road Interchanges

A state employee wishes to see if there is a significant difference in the number of employees at the interchanges of three state toll roads. The data are shown. At $\alpha = 0.05$, can it be concluded that there is a significant difference in the average number of employees at each interchange?

Pennsylvania Turnpike	Greensburg Bypass/ Mon-Fayette Expressway	Beaver Valley Expressway
7	10	1
14	1	12
32	1	1
19	0	9
10	11	1
11	1	11
$\bar{X}_1 = 15.5$	$\bar{X}_2 = 4.0$	$\bar{X}_3 = 5.8$
$s_1^2 = 81.9$	$s_2^2 = 25.6$	$s_3^2 = 29.0$

Source: Pennsylvania Turnpike Commission.

Solution

Step 1 State the hypotheses and identify the claim.

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

H_1: At least one mean is different from the others (claim).

Step 2 Find the critical value. Since $k = 3$, $N = 18$, and $\alpha = 0.05$,

$$\text{d.f.N.} = k - 1 = 3 - 1 = 2$$

$$\text{d.f.D.} = N - k = 18 - 3 = 15$$

The critical value is 3.68.

Step 3 Compute the test value.

a. Find the mean and variance of each sample (these values are shown below the data columns in the example).

b. Find the grand mean.

$$\bar{X}_{GM} = \frac{\Sigma X}{N} = \frac{7 + 14 + 32 + \cdots + 11}{18} = \frac{152}{18} = 8.4$$

c. Find the between-group variance.

$$s_B^2 = \frac{\Sigma n_i(\bar{X}_i - \bar{X}_{GM})^2}{k - 1}$$

$$= \frac{6(15.5 - 8.4)^2 + 6(4 - 8.4)^2 + 6(5.8 - 8.4)^2}{3 - 1}$$

$$= \frac{459.18}{2} = 229.59$$

d. Find the within-group variance.

$$s_W^2 = \frac{\Sigma(n_i - 1)s_i^2}{\Sigma(n_i - 1)}$$

$$= \frac{(6 - 1)(81.9) + (6 - 1)(25.6) + (6 - 1)(29.0)}{(6 - 1) + (6 - 1) + (6 - 1)}$$

$$= \frac{682.5}{15} = 45.5$$

e. Find the F test value.

$$F = \frac{s_B^2}{s_W^2} = \frac{229.59}{45.5} = 5.05$$

Step 4 Make the decision. Since $5.05 > 3.68$, the decision is to reject the null hypothesis.

Step 5 Summarize the results. There is enough evidence to support the claim that there is a difference among the means. The ANOVA summary table for this example is shown in Table 12–3.

Table 12–3 Analysis of Variance Summary Table for Example 12–2

Source	Sum of squares	d.f.	Mean square	F
Between	459.18	2	229.59	5.05
Within	682.5	15	45.5	
Total	1141.68	17		

The steps for computing the F test value for the ANOVA are summarized in this Procedure Table.

Procedure Table

Finding the F Test Value for the Analysis of Variance

Step 1 Find the mean and variance of each sample.

$$(\bar{X}_1, s_1^2), (\bar{X}_2, s_2^2), \ldots, (\bar{X}_k, s_k^2)$$

Step 2 Find the grand mean.

$$\bar{X}_{GM} = \frac{\Sigma X}{N}$$

Step 3 Find the between-group variance.

$$s_B^2 = \frac{\Sigma n_i(\bar{X}_i - \bar{X}_{GM})^2}{k-1}$$

Step 4 Find the within-group variance.

$$s_W^2 = \frac{\Sigma(n_i - 1)s_i^2}{\Sigma(n_i - 1)}$$

Step 5 Find the F test value.

$$F = \frac{s_B^2}{s_W^2}$$

The degrees of freedom are

d.f.N. $= k - 1$

where k is the number of groups, and

d.f.D. $= N - k$

where N is the sum of the sample sizes of the groups

$$N = n_1 + n_2 + \cdots + n_k$$

The P-values for ANOVA are found by using the procedure shown in Section 9–2. For Example 12–2, find the two α values in the tables for the F distribution (Table H), using d.f.N. $= 2$ and d.f.D. $= 15$, where $F = 5.05$ falls between. In this case, 5.05 falls between 4.77 and 6.36, corresponding, respectively, to $\alpha = 0.025$ and $\alpha = 0.01$; hence, $0.01 < P\text{-value} < 0.025$. Since the P-value is between 0.01 and 0.025 and since $P\text{-value} < 0.05$ (the originally chosen value for α), the decision is to reject the null hypothesis. (The P-value obtained from a calculator is 0.021.)

When the null hypothesis is rejected in ANOVA, it only means that at least one mean is different from the others. To locate the difference or differences among the means, it is necessary to use other tests such as the Tukey or the Scheffé test.

Applying the Concepts 12–1

Colors That Make You Smarter

The following set of data values was obtained from a study of people's perceptions on whether the color of a person's clothing is related to how intelligent the person looks. The subjects rated the person's intelligence on a scale of 1 to 10. Group 1 subjects were randomly shown people with clothing in shades of blue and gray. Group 2 subjects were randomly shown people with clothing in shades of brown and yellow. Group 3 subjects were randomly shown people with clothing in shades of pink and orange. The results follow.

Group 1	Group 2	Group 3
8	7	4
7	8	9
7	7	6
7	7	7
8	5	9
8	8	8
6	5	5
8	8	8
8	7	7
7	6	5
7	6	4
8	6	5
8	6	4

1. Use ANOVA to test for any significant differences between the means.
2. What is the purpose of this study?
3. Explain why separate t tests are not accepted in this situation.

See page 668 for the answers.

Exercises 12–1

1. What test is used to compare three or more means?
2. State three reasons why multiple t tests cannot be used to compare three or more means.
3. What are the assumptions for ANOVA?
4. Define between-group variance and within-group variance.
5. What is the F test formula for comparing three or more means?
6. State the hypotheses used in the ANOVA test.
7. When there is no significant difference among three or more means, the value of F will be close to what number?

For Exercises 8 through 19, assume that all variables are normally distributed, that the samples are independent, and that the population variances are equal. Also, for each exercise, perform the following steps.

a. State the hypotheses and identify the claim.
b. Find the critical value.
c. Compute the test value.
d. Make the decision.
e. Summarize the results, and explain where the differences in the means are.

Use the traditional method of hypothesis testing unless otherwise specified.

8. **Sodium Contents of Foods** The amount of sodium (in milligrams) in one serving for a random sample of three different kinds of foods is listed here. At the 0.05 level of significance, is there sufficient evidence to conclude that a difference in mean sodium amounts exists among condiments, cereals, and desserts?

Condiments	Cereals	Desserts
270	260	100
130	220	180
230	290	250
180	290	250
80	200	300
70	320	360
200	140	300
		160

Source: *The Doctor's Pocket Calorie, Fat, and Carbohydrate Counter.*

9. **Hybrid Vehicles** A study was done before the recent surge in gasoline prices to compare the cost to drive 25 miles for different types of hybrid vehicles. The cost of a gallon of gas at the time of the study was approximately \$2.50. Based on the information given below for different models of hybrid cars, trucks, and SUVs, is there sufficient evidence to conclude a

difference in the mean cost to drive 25 miles? Use $\alpha = 0.05$. (The information in this exercise will be used in Exercise 3 in Section 12–2.)

Hybrid cars	Hybrid SUVs	Hybrid trucks
2.10	2.10	3.62
2.70	2.42	3.43
1.67	2.25	
1.67	2.10	
1.30	2.25	

Source: www.fueleconomy.com

10. Healthy Eating Americans appear to be eating healthier. Between 1970 and 2007 the per capita consumption of broccoli increased 1000% from 0.5 to 5.5 pounds. A nutritionist followed a group of people randomly assigned to one of three groups and noted their monthly broccoli intake (in pounds). At $\alpha = 0.05$ is there a difference in means?

Group A	Group B	Group C
2.0	2.0	3.7
1.5	1.5	2.5
0.75	4.0	4.0
1.0	3.0	5.1
1.3	2.5	3.8
3.0	2.0	2.9

Source: *World Almanac.*

11. Lengths of Suspension Bridges The lengths (in feet) of a random sample of suspension bridges in the United States, Europe, and Asia are shown. At $\alpha = 0.05$, is there sufficient evidence to conclude that there is a difference in mean lengths?

United States	Europe	Asia
4260	5238	6529
3500	4626	4543
2300	4347	3668
2000	3300	3379
1850		2874

Source: *New York Times Almanac.*

12. Weight Gain of Athletes A researcher wishes to see whether there is any difference in the weight gains of athletes following one of three special diets. Athletes are randomly assigned to three groups and placed on the diet for 6 weeks. The weight gains (in pounds) are shown here. At $\alpha = 0.05$, can the researcher conclude that there is a difference in the diets?

Diet A	Diet B	Diet C
3	10	8
6	12	3
7	11	2
4	14	5
	8	
	6	

A computer printout for this problem is shown. Use the P-value method and the information in this printout to test the claim. (The information in this exercise will be used in Exercise 4 of Section 12–2.)

Computer Printout for Exercise 12

```
ANALYSIS OF VARIANCE SOURCE TABLE
Source        df    Sum of Squares    Mean Square       F      P-value

Bet Groups     2           101.095         50.548   7.740      0.00797
W/I Groups    11            71.833          6.530

Total         13           172.929

DESCRIPTIVE STATISTICS
Condit         N             Means         St Dev

diet A         4             5.000          1.826
diet B         6            10.167          2.858
diet C         4             4.500          2.646
```

13. Expenditures per Pupil The per-pupil costs (in thousands of dollars) for cyber charter school tuition for school districts in three areas of southwestern Pennsylvania are shown. At $\alpha = 0.05$, is there a difference in the means? If so, give a possible reason for the difference. (The information in this exercise will be used in Exercise 5 of Section 12–2.)

Area I	Area II	Area III
6.2	7.5	5.8
9.3	8.2	6.4
6.8	8.5	5.6
6.1	8.2	7.1
6.7	7.0	3.0
6.9	9.3	3.5

Source: *Tribune-Review.*

14. Cell Phone Bills The average local cell phone monthly bill is $50.07. A random sample of monthly bills from three different providers is listed below. At $\alpha = 0.05$ is there a difference in mean bill amounts among providers?

Provider X	Provider Y	Provider Z
48.20	105.02	59.27
60.59	85.73	65.25
72.50	61.95	70.27
55.62	75.69	42.19
89.47	82.11	52.34

Source: *World Almanac.*

15. Number of Farms The numbers (in thousands) of farms per state found in three sections of the country are listed next. Test the claim at $\alpha = 0.05$ that the mean number of farms is the same across these three geographic divisions.

Eastern third	Middle third	Western third
48	95	29
57	52	40
24	64	40
10	64	68
38		

Source: *New York Times Almanac.*

16. Annual Child Care Costs Annual child care costs for infants are considerably higher than for older children. At $\alpha = 0.05$, can you conclude a difference in mean infant day care costs for different regions of the United States? (Annual costs per infant are given in dollars.) (The information in this exercise will be used in Exercise 6 of Section 12–2.)

New England	Midwest	Southwest
10,390	9,449	7,644
7,592	6,985	9,691
8,755	6,677	5,996
9,464	5,400	5,386
7,328	8,372	

Source: www.naccrra.org (National Association of Child Care Resources and Referral Agencies: "Breaking the Piggy Bank").

17. Microwave Oven Prices A research organization tested microwave ovens. At $\alpha = 0.10$, is there a significant difference in the average prices of the three types of oven?

Watts		
1000	900	800
270	240	180
245	135	155
190	160	200
215	230	120
250	250	140
230	200	180
	200	140
	210	130

A computer printout for this exercise is shown. Use the *P*-value method and the information in this printout to test the claim. (The information in this exercise will be used in Exercise 7 of Section 12–2.)

Computer Printout for Exercise 17

```
ANALYSIS OF VARIANCE SOURCE TABLE
Source       df     Sum of Squares     Mean Square        F        P-value
Bet Groups    2        21729.735        10864.867      10.118      0.00102
W/I Groups   19        20402.083         1073.794
Total        21        42131.818

DESCRIPTIVE STATISTICS
Condit        N          Means           St Dev
 1000         6         233.333           28.23
  900         8         203.125           39.36
  800         8         155.625           28.21
```

18. Calories in Fast-Food Sandwiches Three popular fast-food restaurant franchises specializing in burgers were surveyed to find out the number of calories in their frequently ordered sandwiches. At the 0.05 level of significance can it be concluded that a difference in mean number of calories per burger exists? The information in this exercise will be used for Exercise 8 in Section 12–2.

FF#1	FF#2	FF#3
970	1010	740
880	970	540
840	920	510
710	850	510
	820	

Source: www.fatcalories.com