9.3: The F-Distribution and the F-Ratio

The distribution used for the hypothesis test is a new one. It is called the ***F*-distribution**, invented by George Snedecor but named in honor of Sir Ronald Fisher, an English statistician. The F-statistic is a ratio (a fraction). There are two sets of degrees of freedom; one for the numerator and one for the denominator.

Here are some facts about the *F*-distribution.

1. The curve is not symmetrical but skewed to the right.
2. There is a different curve for each set of degrees of freedom.
3. The F-statistic is greater than or equal to zero.
4. As the degrees of freedom for the numerator and for the denominator get larger, the curve approximates the normal as can be seen in the two figures below. Notice that with more degrees of freedom as shown in the figure below, the curve is more closely approaching the normal distribution, but remember that the F cannot ever be less than zero so the distribution does not have a tail that goes to infinity on the left as the normal distribution does.
5. Other uses for the F-distribution include comparing two variances and two-way Analysis of Variance. Two-Way Analysis is beyond the scope of this chapter.

This graph has an unmarked Y axis and then an X axis that ranges from 0.00 to 4.00. It has three plot lines. The plot line labelled F subscript 1, 5 starts near the top of the Y axis at the extreme left of the graph and drops quickly to near the bottom at 0.50, at which point is slowly decreases in a curved fashion to the 4.00 mark on the X axis. The plot line labelled F subscript 100, 100 remains at Y = 0 for much of its length, except for a distinct peak between 0.50 and 1.50. The peak is a smooth curve that reaches about half way up the Y axis at its peak. The plot line labelled F subscript 5, 10 increases slightly as it progresses from 0.00 to 0.50, after which it peaks and slowly decreases down the remainder of the X axis. The peak only reaches about one fifth up the height of the Y axis.

Figure $9.3.1$

For example, if F follows an F distribution and the number of degrees of freedom for the numerator is four, and the number of degrees of freedom for the denominator is ten, then $F \sim F_{4,10}$.

To calculate the ***F*-ratio**, two estimates of the variance are made.

1. **Variance between samples:** An estimate of σ^2 that is the variance of the sample means multiplied by *n*. (If the samples are different sizes, the variance between samples is thus weighted to account for the different sample sizes.) This variance is also called **variation due to treatment or group,** or **explained variation.** It may also be called **"factor" variance**.

$$MS_{\text{Between}} = \frac{SS_{\text{Between}}}{df_{\text{Between}}} = \frac{n_1(\bar{x}_1 - \bar{x})^2 + \cdots + n_g(\bar{x}_g - \bar{x})^2}{g-1}$$

2. **Variance within samples:** An estimate of σ^2 that is the average of the sample variances. When the sample sizes are different, the variance within samples is weighted. This variance is also called the **variation due to error,** or **unexplained variation.**

$$MS_{\text{Within}} = \frac{SS_{\text{Within}}}{df_{\text{Within}}} = \frac{(n_1 - 1)s_1^2 + \cdots + (n_g - 1)s_g^2}{n-g}$$

- $SS_{between}$ = the **sum of squares** that represents the variation among the different samples
- SS_{within} = the sum of squares that represents the variation within samples that is due to chance.

To find a "sum of squares" means to add together squared quantities that, in some cases, may be weighted. We used sum of squares to calculate the sample variance and the sample standard deviation.

MS in these equations means "**mean square**". $MS_{between}$ is the variance between groups, and MS_{within} is the variance within groups.

The one-way ANOVA test depends on the fact that $MS_{between}$ can be influenced by population differences among means of the several groups. Since MS_{within} compares values of each group to its own group mean, the fact that group means might be different does not affect MS_{within}.

The null hypothesis says that all groups are samples from populations having the same normal distribution. The alternate hypothesis says that at least two of the sample groups come from populations with different normal distributions.

> **Note**
>
> The null hypothesis says that all the group population means are equal. The hypothesis of equal means implies that the populations have the same normal distribution, because it is assumed that the populations are normal and that they have equal variances.

F-Ratio or F-Statistic

$$F_{obs} = \frac{MS_{\text{between}}}{MS_{\text{within}}}$$

If $MS_{between}$ and MS_{within} estimate the same value (following the belief that H_0 is true), then the F-ratio should be approximately equal to one. Mostly, just sampling errors would contribute to variations away from one. As it turns out, $MS_{between}$ consists of the population variance plus a variance produced from the differences between the samples. MS_{within} is an estimate of the population variance. If the null hypothesis is false, $MS_{between}$ will generally be larger than MS_{within}, and the F-ratio will be larger than one. However, if the population effect is small, it is not unlikely that MS_{within} will be larger in a given sample.

To determine the critical value, we have to find $F_{\alpha,df1,df2}$. See Appendix A for the F table. This table has F values for various levels of significance on different pages, as indicated in the first row of each page's table. To find the critical value, choose the page/table with the desired significance level, and follow down and across to find the critical value at the intersection of the two different degrees of freedom. The F distribution has two different degrees of freedom, one associated with the numerator, df_1, and one associated with the denominator, df_2. The degrees of freedom in the numerator is $g-1$, where g is the number of groups; and the degrees of freedom in the denominator is $n-g$, where n is the total sample size across all groups. $F_{\alpha,df1,df2}$ will give the critical value on the upper end of the F distribution.

Data are typically put into a table for easy viewing. One-way ANOVA results are often displayed in this manner by computer software.

Table 9.3.1

Source of variation	Sum of squares (SS)	Degrees of freedom (df)	Mean square (MS)	F
Factor (Between)	SS(Factor)	$g-1$	$MS(Factor) = SS(Factor)/(g-1)$	$\boldsymbol{F = MS(Factor)/MS(E}$
Error (Within)	SS(Error)	$n-g$	$MS(Error) = SS(Error)/(n-g)$	
Total	SS(Total)	$n-1$		

Example 9.3.1

Three different diet plans are to be tested for mean weight loss. The entries in the table are the weight losses for the different plans. The one-way ANOVA results are shown in the table below.

Table 9.3.2

Plan 1: $n_1 = 4$	Plan 2: $n_2 = 3$	Plan 3: $n_3 = 3$
5	3.5	8
4	7	4
4	4.5	3
3		

$H_0: \mu_1 = \mu_2 = \mu_3$
H_a: At least two of the means (μ_1, μ_2, μ_3) are not equal.
Means: $\bar{x}_1 = \frac{16}{4} = 4, \bar{x}_2 = \frac{15}{3} = 5, \bar{x}_3 = \frac{15}{3} = 5, \bar{x} = \frac{46}{10} = 4.6$
Variances: $s_1^2 = 0.67, s_2^2 = 3.16, s_3^2 = 7$
$SS_{Between} = 4(4-4.6)^2 + 3(5-4.6)^2 + 3(5-4.6)^2 = 2.4$
$SS_{Within} = (4-1)0.67 + (3-1)3.16 + (3-1)7 = 22.33$
$df_{Between} = 3-1 = 2$
$df_{Within} = 10-3 = 7$
$MS_{Between} = \frac{SS_{Between}}{df_{Between}} = \frac{2.4}{2} = 1.2$
$MS_{Within} = \frac{SS_{Within}}{df_{Within}} = \frac{22.33}{7} = 3.19$
$F_{obs} = \frac{MS_{Between}}{MS_{Within}} = \frac{1.2}{3.19} = 0.38$

Table 9.3.3

Source of variation	Sum of squares (SS)	Degrees of freedom (df)	Mean square (MS)	F
Factor (Between)	2.4	2	1.2	0.38

Source of variation	Sum of squares (SS)	Degrees of freedom (df)	Mean square (MS)	F
Error (Within)	22.33	7	3.19	
Total	2.4 + 22.33 = 24.73	2 + 7 = 9		

For $F_{\alpha=.05,2,7}$, $F_{critical}$ = 4.74. F_{obs} does not exceed this $F_{critical}$, so we cannot reject H_0; we don't have sufficient evidence to say that the three different diet plans result in different average amounts of weight loss.

Exercise 9.3.1

As part of an experiment to see how different types of soil cover would affect slicing tomato production, Marist College students grew tomato plants under different soil cover conditions. Groups of three plants each had one of the following treatments

- bare soil
- a commercial ground cover
- black plastic
- straw
- compost

All plants grew under the same conditions and were the same variety. Students recorded the weight (in grams) of tomatoes produced by each of the *n* = 15 plants:

Table 9.3.4

Bare: $n_1 = 3$	Ground Cover: $n_2 = 3$	Plastic: $n_3 = 3$	Straw: $n_4 = 3$	Compost: $n_5 = 3$
2,625	5,348	6,583	7,285	6,277
2,997	5,682	8,560	6,897	7,818
4,915	5,482	3,830	9,230	8,677

Create the one-way ANOVA table.

The one-way ANOVA hypothesis test is always right-tailed because larger F-values are way out in the right tail of the F-distribution curve.

Example 9.3.2

Let's return to the slicing tomato exercise we were working on in the previous exercise. The means of the tomato yields under the five mulching conditions are represented by $\mu_1, \mu_2, \mu_3, \mu_4, \mu_5$. We will conduct a hypothesis test to determine if all means are the same or at least one is different. Using a significance level of 5%, test the null hypothesis that there is no difference in mean yields among the five groups against the alternative hypothesis that at least one mean is different from the rest.

Answer

The null and alternative hypotheses are:

$H_0 : \mu_1 = \mu_2 = \mu_3 = \mu_4 = \mu_5$

H_a : at least two of the population means differ

The one-way ANOVA results are shown in the table below.

Table 9.3.5

Source of variation	Sum of squares (SS)	Degrees of freedom (df)	Mean square (MS)	F
Factor (Between)	36,648,561	$5-1=4$	$\frac{36{,}648{,}561}{4} = 9,162,140$	$\frac{9{,}162{,}140}{2{,}044{,}672.6} = 4.481$
Error (Within)	20,446,726	$15-5=10$	$\frac{20{,}446{,}726}{10} = 2,044,672.6$	
Total	57,095,287	$15-1=14$		

Distribution for the test: $F_{\alpha=.05,4,10}$

$df_{num} = 5 - 1 = 4$

$df_{denom} = 15 - 5 = 10$

$F_{critical} = 3.48$

Test statistic: $F_{obs} = 4.481$

This graph shows a nonsymmetrical F distribution curve. The horizontal axis extends from 0 - 5, and the vertical axis ranges from 0 - 0.7. The curve is strongly skewed to the right.

Figure $9.3.2$

Make a decision: Since $F_{obs} > F_{critical}$, we reject H_0.

Conclusion: At the 5% significance level, we have reasonably strong evidence that differences in mean yields for slicing tomato plants grown under different mulching conditions are unlikely to be due to chance alone. We may conclude that at least one of mulches led to a different mean yield than the others.

Exercise $9.3.2$

MRSA, or *Staphylococcus aureus*, can cause a serious bacterial infections in hospital patients. The table below shows various colony counts from different patients who may or may not have MRSA. The data from the table is plotted in the figure below.

Table 9.3.6

Conc = 0.6	Conc = 0.8	Conc = 1.0	Conc = 1.2	Conc = 1.4
9	16	22	30	27
66	93	147	199	168
98	82	120	148	132

Plot of the data for the different concentrations:

This graph is a scatterplot for the data provided. The horizontal axis is labeled 'Colony counts' and extends from 0 - 200. The vertical axis is labeled 'Tryptone concentrations' and extends from 0.6 - 1.4.

Figure $9.3.3$

Test whether the mean number of colonies are the same or are different. Construct the ANOVA table, find the *p*-value, and state your conclusion. Use a 5% significance level.

Example $9.3.3$

Four sororities took a random sample of their members regarding their grade means for the past term. The results are shown in the table below.

Table 9.3.7

Sorority 1	Sorority 2	Sorority 3	Sorority 4
2.17	2.63	2.63	3.79
1.85	1.77	3.78	3.45
2.83	3.25	4.00	3.08
1.69	1.86	2.55	2.26
3.33	2.21	2.45	3.18

Using a significance level of 1%, is there a difference in mean grades among the sororities?

Answer

> **Note**
>
> This is an example of a **balanced design**, because each group (i.e., sorority) has the same number of observations.

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

H_a: Not all of the means $\mu_1, \mu_2, \mu_3, \mu_4$ are equal.

Distribution for the test: $F_{\alpha=.01,3,16}$

where $g = 4$ groups and $n = 20$ samples in total

$df_{num} = g - 1 = 4 - 1 = 3$

$df_{denom} = n - g = 20 - 4 = 16$

$F_{critical} = 5.29$

Calculate the test statistic: $F_{obs} = 2.23$

Graph:

This graph shows a nonsymmetrical F distribution curve with values of 0 and 2.23 on the x-axis representing the test statistic of sorority grade averages. The curve is slightly skewed to the right, but is approximately normal. A vertical upward line extends from 2.23 to the curve and the area to the right of this is shaded to represent the p-value.

Figure $9.3.4$

Make a decision: Since $F_{obs} < F_{critical}$, we reject H_0.

Conclusion: There is not sufficient evidence to conclude that there is a difference among the mean grades for the sororities.

Exercise $9.3.3$

Four sports teams took a random sample of players regarding their GPAs for the last year. The results are shown here in the table.

Table 9.3.8

Basketball	Baseball	Hockey	Lacrosse
3.6	2.1	4.0	2.0
2.9	2.6	2.0	3.6
2.5	3.9	2.6	3.9
3.3	3.1	3.2	2.7
3.8	3.4	3.2	2.5

Use a significance level of 5%, and determine if there is a difference in GPA among the teams.

Example $9.3.4$

A fourth grade class is studying the environment. One of the assignments is to grow bean plants in different soils. Tommy chose to grow his bean plants in soil found outside his classroom mixed with dryer lint. Tara chose to grow her bean plants in potting soil bought at the local nursery. Nick chose to grow his bean plants in soil from his mother's garden. No chemicals were used on the plants, only water. They were grown inside the classroom next to a large window. Each child grew five plants. At the end of the growing period, each plant was measured, producing the data (in inches) in the table shown here.

Table 9.3.9

Tommy's plants	Tara's plants	Nick's plants
24	25	23
21	31	27
23	23	22
30	20	30

Tommy's plants	Tara's plants	Nick's plants
23	28	20

Does it appear that the three media in which the bean plants were grown produce the same mean height? Test at a 5% level of significance.

Answer

$H_0 : \mu_{Tommy} = \mu_{Tara} = \mu_{Nick}$
H_a: At least two of the means (μ_{Tommy} , μ_{Tara} , μ_{Nick}) are not equal.
Means: $\bar{x}_{Tommy} = \frac{121}{5} = 24.2, \bar{x}_{Tara} = \frac{127}{5} = 25.4, \bar{x}_{Nick} = \frac{122}{5} = 24.4, \bar{x} = \frac{370}{15} = 24.7$
Variances: $s^2_{Tommy} = 11.7, s^2_{Tara} = 18.3, s^2_{Nick} = 16.3$
$SS_{Between} = 5(24.2 - 24.7)^2 + 5(25.4 - 24.7)^2 + 5(24.4 - 24.7)^2 = 4.15$
$SS_{Within} = (5-1)11.7 + (5-1)18.3 + (5-1)16.3 = 185.2$
$df_{Between} = 3 - 1 = 2$
$df_{Within} = 15 - 3 = 12$
$MS_{Between} = \frac{SS_{Between}}{df_{Between}} = \frac{4.15}{2} = 2.075$
$MS_{Within} = \frac{SS_{Within}}{df_{Within}} = \frac{185.2}{12} = 15.43$
$F_{obs} = \frac{MS_{Between}}{MS_{Within}} = \frac{2.075}{15.43} = 0.13$

Table 9.3.10

Source of variation	Sum of squares (SS)	Degrees of freedom (df)	Mean square (MS)	F
Factor (Between)	4.15	2	2.075	0.13
Error (Within)	185.2	12	15.43	
Total	189.35	14		

For $F_{\alpha=.05,2,12}$, $F_{critical}$ = 3.89. F_{obs} does not exceed this $F_{critical}$, so we cannot reject H_0. With a 5% level of significance, from this sample data, the evidence is not sufficient to conclude that the mean heights of the bean plants are different.