Dr. Saeed A. Dobbah Alghamdi

Department of Statistics
Faculty of Sciences
Building 90, Office 26F41
King Abdulaziz University
http://saalghamdy.kau.edu.sa

Dr. Saeed's Website

Statistics
for
Business & Economics

David R. Anderson
Dennis J. Sweeney
Thomas A. Williams

Analysis of Variance

Chapter 11

Learning Objectives

LO1 List the characteristics of the *F* distribution.

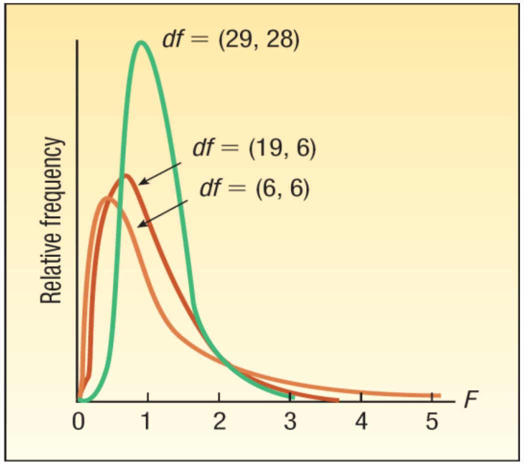
LO2 Perform a test of hypothesis to determine the equality of two variances.

LO3 Describe the ANOVA approach for testing difference in means.

LO4 Conduct a test of hypothesis among three or more treatment means.

LO5 Conduct a pairwise tests for the difference in treatment means.

The *F* Distribution



Characteristics of the F Distribution

1. There is a “family” of *F* Distributions. A particular member of the family is determined by two parameters: the degrees of freedom in the numerator and in the denominator.
2. The *F* distribution is continuous
3. *F* cannot be negative.
4. The *F* distribution is positively skewed.
5. It is asymptotic. As $F \rightarrow \infty$ the curve approaches the *X*-axis but never touches it.

Uses of the F Distribution

- ☐ test whether two samples are from populations having equal variances.
- ☐ to compare several population means simultaneously. The simultaneous comparison of several population means is called analysis of variance(ANOVA).

Assumption:

In both of the uses above, the populations must follow a normal distribution, and the data must be at least interval-scale.

Comparing Two Population Variances

In some statistical applications we may want to compare the variances in product quality resulting from two different production processes, the variances in assembly times for two assembly methods, or the variances in temperatures for two heating devices.

$H_0: \sigma_1^2 = \sigma_2^2$ $\qquad$ $H_1: \sigma_1^2 \neq \sigma_2^2$

TEST STATISTIC FOR HYPOTHESIS TESTS ABOUT POPULATION VARIANCES WITH $\sigma_1^2 = \sigma_2^2$

$$F = \frac{s_1^2}{s_2^2}$$

Denoting the population with the larger sample variance as population 1, the test statistic has an F distribution with $n_1 - 1$ degrees of freedom for the numerator and $n_2 - 1$ degrees of freedom for the denominator.

Test for Equal Variances

Example

The variance in a production process is an important measure of the quality of the process. A large variance often signals an opportunity for improvement in the process by finding ways to reduce the process variance. Conduct a statistical test to determine whether there is a significant difference between the variances in the bag weights for two machines. Use a .05 level of significance. What is your conclusion?

Machine 1	2.95	3.45	3.50	3.75	3.48	3.26	3.33	3.20
	3.16	3.20	3.22	3.38	3.90	3.36	3.25	3.28
	3.20	3.22	2.98	3.45	3.70	3.34	3.18	3.35
	3.12							
Machine 2	3.22	3.30	3.34	3.28	3.29	3.25	3.30	3.27
	3.38	3.34	3.35	3.19	3.35	3.05	3.36	3.28
	3.30	3.28	3.30	3.20	3.16	3.33		

Step 1: The hypotheses are:

$H_0: \sigma_1^2 = \sigma_2^2$ $H_1: \sigma_1^2 \neq \sigma_2^2$

Step 2: The significance level is 0.05.

Step 3: The test statistic is the F distribution.

$$F = \frac{S_1^2}{S_2^2} = \mathbf{8.28}$$

Step 4: State the decision rule.

Reject H_0 if *p-value* ≤ 0.05

Step 5: make a decision

F-test for equality of variance

0.0489	variance: machine 1
0.0059	variance: machine 2
8.28	F
7.22E-06	p-value

The p-value is 0.0000 which is less than 0.05 indicating the **rejection of the null hypothesis**. Hence, **there is a difference in the variation** of bag weights for the two machines.

Comparing Means of Three or More Populations

The F distribution is also used for testing the equality of three or more population means. This test is Known as Analysis of Variance (ANOVA) test.

Assumptions for ANOVA test:

- For each population, the response variable is normally distributed.
- The variance of the response variable is the same for all of the populations.
- The observations must be independent.

Hypothesis:

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

H_1: Not all populations means are equal

Reject H_0 if *p-value* $< \alpha$

FIGURE 13.2 SAMPLING DISTRIBUTION OF $\bar{x}$ GIVEN H_0 IS TRUE

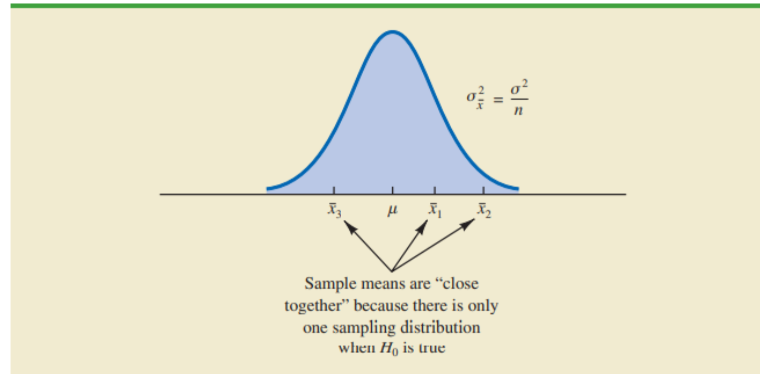
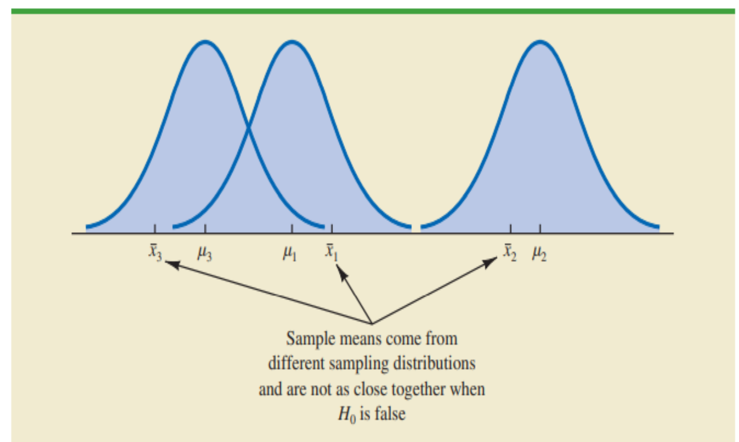


FIGURE 13.3 SAMPLING DISTRIBUTIONS OF $\bar{x}$ GIVEN H_0 IS FALSE



Comparing Means of Two or More Populations-Example

Example

NCP manufactures printers and fax machines at plants located in Atlanta, Dallas, and Seattle. To measure how much employees at these plants know about quality management, a random sample of six employees was selected from each plant and the employees selected were given a quality awareness examination

Managers want to use these data to test the hypothesis that the mean examination score is the same for all three plants at 0.05 significance level

Step 1: State the null and alternate hypotheses.

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

H_a: Not all population means are equal

Step 2: State the level of significance.

Use 0.05 significance level

Plant 1 Atlanta	Plant 2 Dallas	Plant 3 Seattle
85	71	59
75	75	64
82	73	62
76	74	69
71	69	75
85	82	67

Step 3: Find the appropriate test statistic.

Use the F statistic

Comparing Means of Two or More Populations-Example

Step 4: State the decision rule.

Reject H_0 if p-value ≤ 0.05

Step 5: Make a decision

Source	SS	df	MS	F	p-value
Treatment	516.00	2	258.000	9.00	.0027
Error	430.00	15	28.667		
Total	946.00	17			

Since the p-value 0.0027 is less than the significance level 0.05, then reject the **null hypothesis**. we can **only conclude there is a difference in the treatment means**. We cannot determine which treatment groups differ or how many treatment groups differ.

Post Hoc Test for the Differences Between Means

When we reject the null hypothesis that the means are equal, we may want to know which treatment means differ. One of procedures is through the use post hoc tests to tell us which groups differ from the rest. There are a number of tests which can be used, but we will concentrate on the use of pairwise t-test which is one of the MEGASTAT output when using the one way ANOVA.

Example

From the previous example, which treatments means are different at 0.05 significance level?

p-values for pairwise t-tests				
		Seattle	Dalas	Atlanta
		66.0	74.0	79.0
Seattle	66.0			
Dalas	74.0	.0206		
Atlanta	79.0	.0008	.1266	

Note that there is a significant difference in the treatments means between Seattle and Dalas and between Atlanta and Seattle at 0.05 significance level. Thus, the quality awareness in Seattle should be reviewed.