Chapter 13: F Distribution and One-Way ANOVA

13.1: One-Way Analysis of Variance

In previous sections, we compared two population means. We will now extend the concept of comparing two population means to comparing three or more population means. The procedure for doing this is called **one-way analysis of variance**, or **ANOVA** for short.

If k is the number of populations, the null and alternative hypotheses for comparing population means can be written as

$$H_0 : \mu_1 = \mu_2 = \ldots = \mu_k$$
$$H_a : \text{The population means are not all equal}$$

We follow the same five steps for hypothesis testing that we have used in other chapters, but our test statistic is no longer a z or a t statistic. The appropriate test statistic is called an **F-statistic**, the general procedure is called *one-way analysis of variance*, and the significance test is called an **F-test**:

$$F = \frac{\text{Variation BETWEEN groups}}{\text{Variation WITHIN groups}}$$

The variation between sample means is 0 if all k of the sample means are exactly equal and gets larger the more spread out they are.

To find the p-value, which is the probability that the computed F-statistic would be as large as it is (or larger) if the null hypothesis is true, a probability distribution called the **F-distribution** is used.

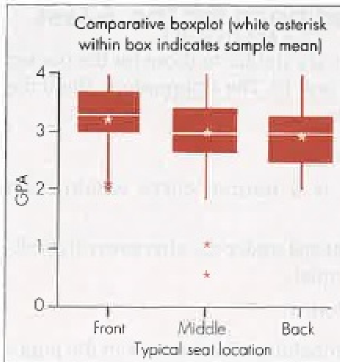
Example 1. Is it true that the best students sit in the front of a classroom, or is that a false stereotype? In surveys done in two statstics classes at the University of California at Davis, students reported their grade point averages and also answered the question, "Where do you typically sit in a classroom (front, middle, back)?" In all, 384 students gave valid responses to both questions, and among these students, 88 said that they typically sit in the front, 218 said they typically sit in the middle, and 78 said they typically sit in the back.

The figure below shows a boxplot comparing the GPAs for the three seat locations and Minitab analysis of variance results that can be used to test whether or not the mean GPAs are the same for the three locations. In the boxplot, we see that students who preferred sitting in the front generally had slightly higher GPAs than the others; toward the bottom of the computer output, the sample means are given as Front = 3.2029, Middle = 2.9853, Back = 2.9194.

The analysis of variance results in the table can be used to test the following:

$$H_0 : \mu_1 = \mu_2 = \mu_3$$
$$H_a : \text{The three means are not all equal}$$

where μ_1, μ_2, and μ_3 are the population GPAs for the populations of students who typically sit in the front, middle, and back of the classroom, respectively. The p-value for the F-test is under "P" in the rightmost column of the table titled "Analysis of Variance for GPA," and its value is 0.001. With a p-value this small, we can reject the null hypothesis and thus conclude that there are differences among the means in the populations represented by the samples.



Analysis of Variance for GPA

Source	DF	SS	MS	F	P
Seat	2	3.994	1.997	6.69	**0.001**
Error	381	113.775	0.299		
Total	383	117.769			

Level	N	Mean	StDev
Front	88	3.2029	0.5491
Middle	218	2.9853	0.5577
Back	78	2.9194	0.5105

Pooled StDev = 0.5465

```
Individual 95% CIs For Mean Based on Pooled StDev
----+---------+---------+---------+--
                   (-------*-------)
       (----*----)
 (-------*-------)
----+---------+---------+---------+--
  2.85      3.00      3.15      3.30
```

Note that the output also shows individual 95% confidence intervals for the three population means. The location of the interval for the Front" mean does not overlap with the other two intervals, which indicates a significant difference between the mean GPA for the front-row sitters and the mean GPA for the other students. It is not clear whether there is a significant difference between the "Middle" and "Back" groups, since their confidence intervals overlap.

13.2: The F Distribution and the F-Ratio

Notation for Summary Statistics

Useful notation for summarizing statistics from the observed samples is

k = number of groups
$\bar{x}_i$, s_i, and n_i are the mean, standard deviation, and sample size for the i-th sample group
N = total sample size $= n_1 + n_2 + \ldots + n_k$

The Analysis of Variance Table

A fundamental concept in one-way analysis of variance is that the variation among the data values in the overall sample can be separated into (1) differences between group means, and (2) natural variation among observations within a group.

The calculations and theory for ANOVA stem from the fact that for a particular way of measuring variation, the sum of the variation between group means and the variation among observations within groups equals the total variation:

$$\text{Total variation} \;=\; \text{Variation between group means} \;+\; \text{Variation within groups}$$

The F-statistic measures the relative size of the variation between group means and the natural variation within groups.

Measuring Variation BETWEEN Group Means

The variation between group means is measured with a weighted sum of squared differences between the sample means ($\bar{x}_i$) and $\bar{x}$, the overall mean of all data.

Each squared difference is multiplied by the appropriate group sample size, n_i in this sum.

This quantity is called **sum of squares for groups** or **SS Groups**:

$$\text{SS Groups } = n_1(\bar{x}_1 - \bar{x})^2 + n_2(\bar{x}_2 - \bar{x})^2 + \cdots + n_k(\bar{x}_k - \bar{x})^2 = \sum_{\text{groups}} n_i(\bar{x}_i - \bar{x})^2$$

The numerator of the F-statistic for comparing means is called the **mean square for groups** or **MS Groups**:

$$\text{MS Groups } = \frac{\text{SS Groups}}{k-1} = \frac{SST}{k-1}$$

Measuring Variation WITHIN Groups

To measure variation among individuals within groups, first find the sum of squared deviations between data values and the group mean separately for each group.

Then, add these quantities, this is called the **sum of squared errors** or **SSE**:

$$\text{SS Error } = (n_1 - 1)s_1^2 + (n_2 - 1)s_2^2 + \cdots + (n_k - 1)s_k^2 = \sum_{\text{groups}} (n_i - 1)s_i^2$$

The denominator of the F-statistic is called the **mean square error** or **MSE**:

$$\text{MSE } = \frac{\text{SSE}}{N - k}$$

Note that MSE is just a weighted average of the sample variances for the k groups. In fact, if all n_i are equal, MSE is simply the average of the k sample variances.

$$F = \frac{\text{MS Groups}}{\text{MSE}} = \frac{MST}{MSE} = \frac{\text{variation BETWEEN group}}{\text{variation WITHIN groups}}$$

More Notation

- $F \sim F(df_{\text{num}}, df_{\text{den}})$
  - where $df_{\text{num}} = k - 1$
  - and $df_{\text{den}} = N - k$

Determining the p-Value

The p-value for the F-test is the area to the right of the value of the F-statistic, under and F-distribution with $k - 1$ and $N - k$ degrees of freedom:

$$df = (k - 1, N - k)$$

The F-Table gives **critical values** for two different levels of significance. If the observed F-statistic is greater than or equal to the critical value for a particular level of significance, the result is statistically significant at that level.

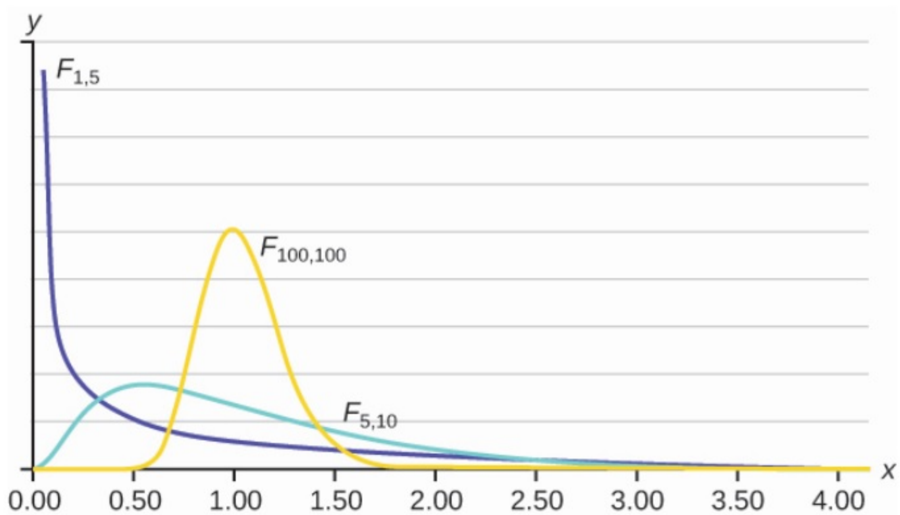
To use the table, first determine values for the numerator degrees of freedom (these are the column headings) and the denominator degrees of freedom (these are the row headings).

13.3: Facts About the F Distribution

The Family of F-Distributions

An F-**distribution** is used to find the p-value for an ANOVA F-test.

1. The F distribution is not symmetric. It is skewed right.
2. Values of the F distribution cannot be negative (must be greater than or equal to 0).
3. The exact shape of the F distribution depends on the two different degrees of freedom.
   (a) They are always given in the order of *numerator df*, *denomintor df*.
   (b) The numerator df $= k - 1$, where k is the number of groups.
   (c) The denomintor df $= N - k$, where N is the total sample size.
4. As the degrees of freedom for the numerator and for the denominator get larger, the curve approximates the normal.
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.



13.5: One-Way ANOVA

The Six Steps for a Two-Sample t-Test

Step 1: Determine the Null and Alternative Hypotheses

$H_0 : \mu_1 = \mu_2 = \ldots = \mu_k$ versus H_a : The μ's are not all equal

Step 2: Verify That The Conditions Are Met And State the Level of Significance

Conditions Required for a Two-Sample t-Test To Be Valid

1. The populations have distributions that are approximately normal. This is a loose requirement because the method works well unless a population has a distribution that is very far from normal.
2. The populations have the same variance σ^2. This is a loose requirement because the method workds well unless the population variances differ by large amounts.
3. The samples are representative of the population.
4. The samples are independent of each other.
5. The different samples are from populations that are categorized in only one way.

Step 3: Summarize the Data into an Appropriate Test Statistic

$$F = \frac{\text{MS Groups}}{\text{MSE}} = \frac{\text{variation BETWEEN group}}{\text{variation WITHIN groups}}$$

Step 4: Find the p-Value OR determine the critical value

Using the F-distribution with numerator $df = k - 1$ and denominator $df = N - k$, the p-value is the area in the tail to the right of the test statistic F.

Step 5: Make a decision based on either the p-value or the rejection region

- If p-value $< \alpha$, Reject H_0
- If the test statistic is in the shaded region (rejection region), reject H_0

Step 6: State your conclusion in terms of the problem

Interpret the conclusion in context of the situation. We should also consider the manner in which the data were collected.

In ANOVA, when the null is rejected, these steps are usually followed by a multiple comparison procedure to determine which group means are significantly different from each other.

Example 2. Compute the F-test statistic for the data below.

x_1	x_2	x_3
4	7	10
5	8	10
6	9	11
6	7	11
4	9	13

The output for the above data is provided below.

Source of Variation	Sum of Squares	Degrees of Freedom	Mean Squares	F-Test Statistic
Treatment	90	2	45	
Error	14	12	1.1667	
Total	104	14		

(a) State your hypotheses.

(b) What is the sum of squares due to treatments, SST, and the sume of squares due to error, SSE?

(c) What are the mean squares, MST and MSE?

(d) Compute the F-test statistic.

(e) What is the critical value?

(f) What is your decision?

Example 3. Fill in the ANOVA table.

Source of Variation	Sum of Squares	Degrees of Freedom	Mean Squares	F-Test Statistic
Treatment	387	2		
Error	8042	27		
Total				

Example 4. Fill in the ANOVA table.

Source of Variation	Sum of Squares	Degrees of Freedom	Mean Squares	F-Test Statistic
Treatment	2814			
Error		36		
Total	7729	39		

Multiple Comparisons

The term **multiple comparisons** is used when two or more comparisons are made to examine the specific pattern of differences among means.

The most commonly analyzed set of multiple comparisons is the set of all **pairwise comparisons** among populations means.

It is tempting to test the null hypothesis $H_0 : \mu_1 = \mu_2 = \mu_3$ by comparing the population means two at a time using previously discussed techniques. If we proceeded this way, we would need to test three different hypotheses:

$$\begin{array}{lllll} H_0 : \mu_1 = \mu_2 & \text{and} & H_0 : \mu_1 = \mu_3 & \text{and} & H_0 : \mu_2 = \mu_3 \\ H_1 : \mu_1 \neq \mu_2 & & H_1 : \mu_1 \neq \mu_3 & & H_1 : \mu_2 \neq \mu_3 \end{array}$$

Each test would have a probability of a Type I error of $\alpha = 0.05$, thus, each test would have a 95% probability of rejecting the null hypothesis when the alternative hypothesis of no difference is true. The probability that all three tests correctly reject the null hypothesis is $0.95^3 = 0.86$ (assuming the tests are independent). There is a $1 - 0.95^3 = 1 - 0.86 = 0.14$, or 14% probability that at least one test will lead to an incorrect rejection of H_0. A 14% probability of a Type I error is much higher than the desired 5% probability. As the number of populations to be compared increases, so does the probability of making a Type I error using multiple t-tests for a given value of α.

We could try adjusting the individual α-values so that the overall probability of a Type I error is the desired value.

However by decreasing the probability of making a Type I error in each test, we increase the probability of making a Type II error for those tests. So, conducting multiple t-tests ultimately leads to an increased chance of making a mistake no matter what level of α we use in the individual tests.

Bonferroni Multiple Comparison Test test the groups in pairs to determine which group(s) is/are significantly different from the others.

The test statistic is:

$$t = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\text{MSE}\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}}$$

where MSE comes from the ANOVA output, and degrees of freedom: $df = N - k$.

Example 5. The data in the table represent the number of pods on a random sample of soybean plants for various plot types. An agricultural researcher wants to determine if the mean numbers of pods for each plot type are equal.

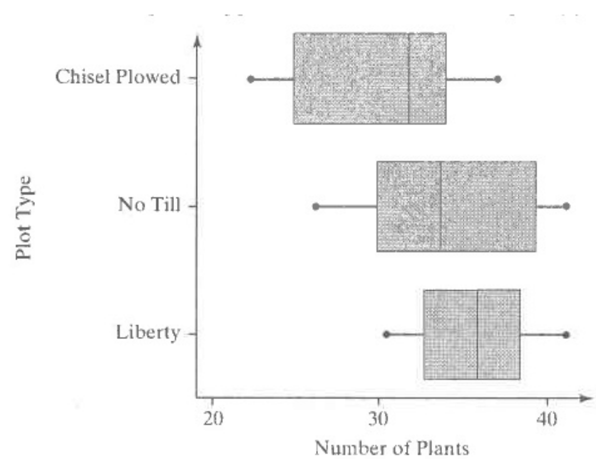
Plot Type	Pods								
Liberty	32	31	36	35	41	34	39	37	38
No till	34	30	31	27	40	33	37	42	39
Chisel Plowed	34	37	24	23	32	33	27	34	30

(a) Write the null and alternative hypotheses.

(b) State the requirements that must be satisfied to use the one-way ANOVA procedure.

(c) Use the following output to test the hypothesis of equal means at the $\alpha = 0.05$ level of significance. Support your decision from part (c), by determining both the p-value and critical value.

Source	DF	SS	MS	F	P
Factor	2	149.0	74.5	3.77	0.038
Error	24	474.7	19.8		
Total	26	623.6			



(d) Shown are side-by-side boxplots of each type of plot. Do these boxplots support the results obtained in part (c)?

(e) Determine the sample means of each group.

(e) Use Bonferroni's Multiple Comparison Tests to further determine which group(s) is/are different from the others.

Example 6. A stock analyst wondered whether the mean rate of return of financial, energy, and utility stocks differed over the past 5 years. He obtained a simple random sample of eight companies from each of the three sectors and obtained the 5-year rates of return shown in the following table (in percent):

Financial	Energy	Utilities
10.76	12.72	11.88
15.05	13.91	5.86
17.01	6.43	13.46
5.07	11.19	9.90
19.50	18.79	3.95
8.16	20.73	3.44
10.38	9.60	7.11
6.75	17.40	15.70

(a) State the null and alternative hypotheses.

(b) Are the mean rates of return different at the $\alpha = 0.05$ level of significance? Use the p-value and the critical value method.

(c) Conduct Bonferroni's Multiple Comparison to confirm your decision in part (b).

SUMMARY

Groups	*Count*	*Sum*	*Average*	*Variance*
Financial	8	92.68	11.585	26.25283
Energy	8	110.77	13.84625	23.68791
Utilities	8	71.3	8.9125	20.52465

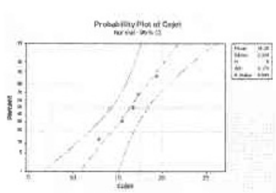
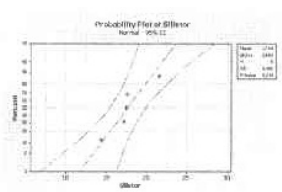
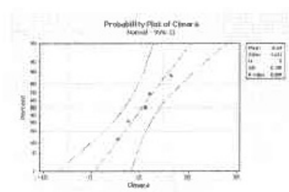
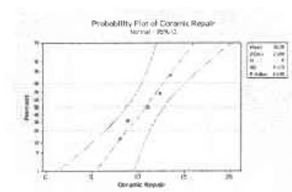
ANOVA

Source of Variation	*SS*	*df*	*MS*	*F*	*P-value*	*F crit*
Between Groups	97.593058	2	48.79653	2.077468	0.150231	3.4668
Within Groups	493.25774	21	23.48846			
Total	590.8508	23				

Example 7. Prosthodentists specialize in the restoration of oral function. Since repairing chipped veneers is less time consuming and less costly than complete restoration, a researcher wanted to determine the effect of different repair kits on shear bond strength for repairs of chipped porcelain veneer in fixed prosthodontics. He randomly divided 20 porcelain specimens into four treatment groups: the Cojet system, the Silistor system, the Cimara system, and the Ceramic Repair system. At the conclusion of the study, shear bond strength (in megapascals, MPa) was measured according to ISO 10477. The results in the table below are based on the results of the study. Verify that the requirements are satisfied and perform a one-way ANOVA test.

Cojet	Silistor	Cimara	Ceramic Repair
15.4	17.2	5.5	11.0
12.9	14.3	7.7	12.4
17.2	17.6	12.2	13.5
16.6	21.6	11.4	8.9
19.3	17.5	16.4	8.1

Q-Q plots and descriptive statistics are provided for each group.

Cojet		*Silistor*		*Cimara*		*Ceramic*	
Mean	16.28	Mean	17.64	Mean	10.64	Mean	10.78
Standard Error	1.05517771	Standard Error	1.163013328	Standard Error	1.887485099	Standard Error	1.019509686
Median	16.6	Median	17.5	Median	11.4	Median	11
Mode	#N/A	Mode	#N/A	Mode	#N/A	Mode	#N/A
Standard Deviation	2.359449088	Standard Deviation	2.600576859	Standard Deviation	4.220544989	Standard Deviation	2.279692962
Sample Variance	5.567	Sample Variance	6.763	Sample Variance	17.813	Sample Variance	5.197
Kurtosis	0.672751477	Kurtosis	2.122777805	Kurtosis	-0.594946939	Kurtosis	-2.182222474
Skewness	-0.34706369	Skewness	0.586381348	Skewness	0.189152602	Skewness	-0.053023557
Range	6.4	Range	7.3	Range	10.9	Range	5.4
Minimum	12.9	Minimum	14.3	Minimum	5.5	Minimum	8.1
Maximum	19.3	Maximum	21.6	Maximum	16.4	Maximum	13.5
Sum	81.4	Sum	88.2	Sum	53.2	Sum	53.9
Count	5	Count	5	Count	5	Count	5

ANOVA

Source of Variation	*SS*	*df*	*MS*	*F*	*P-value*	*F crit*
Between Groups	199.9855	3	66.661833	7.545199		3.238872
Within Groups	141.36	16	8.835			
Total	341.3455	19				

(a) State the hypotheses.

(b) Have the conditions been met to perform the Analysis of Variance test?

(c) Determine the test statistic and make a decision.

(d) Conduct Bonferroni's Multiple Comparison to confirm your decision in part (c).