Lecture 12: Chapter 12

C C Moxley

UAB Mathematics

12 November 15

§12.1 ANOVA (Analysis of Variance) Tests

In Chapter 9, we tested to see if two population means were equal.

§12.1 ANOVA (Analysis of Variance) Tests

In Chapter 9, we tested to see if two population means were equal. In this chapter, we will test to see if three or more population samples are the same using a **one-way** ANOVA test. We call it "one-way" because we separate our populations into groups based on one characteristic.

§12.1 ANOVA (Analysis of Variance) Tests

In Chapter 9, we tested to see if two population means were equal. In this chapter, we will test to see if three or more population samples are the same using a **one-way** ANOVA test. We call it "one-way" because we separate our populations into groups based on one characteristic. This is often called an "analysis of variance" but this refers to the method of testing, not the thing which we are testing - which is means not variances.

§12.1 ANOVA (Analysis of Variance) Tests

In Chapter 9, we tested to see if two population means were equal. In this chapter, we will test to see if three or more population samples are the same using a **one-way** ANOVA test. We call it "one-way" because we separate our populations into groups based on one characteristic. This is often called an "analysis of variance" but this refers to the method of testing, not the thing which we are testing - which is means not variances. We will also compare the population means of populations separated into categories based on two characteristics (such as weight and body temperature).

§12.2 One-Way ANOVA Test

Definition (One-Way ANOVA Test)

A one-way ANOVA test is a method of testing the equality of three or more population means by analyzing sample variances. One-way analysis of variance is used with data categorized with one **factor**, so there is one characteristic used to separate the sample data into the different categories.

§12.2 One-Way ANOVA Test

Definition (One-Way ANOVA Test)

A one-way ANOVA test is a method of testing the equality of three or more population means by analyzing sample variances. One-way analysis of variance is used with data categorized with one **factor**, so there is one characteristic used to separate the sample data into the different categories.

- Populations must be approximately normally distributed.

§12.2 One-Way ANOVA Test

Definition (One-Way ANOVA Test)

A one-way ANOVA test is a method of testing the equality of three or more population means by analyzing sample variances. One-way analysis of variance is used with data categorized with one **factor**, so there is one characteristic used to separate the sample data into the different categories.

- Populations must be approximately normally distributed.
- They must have the same variance, roughly. (Or at least the sample sizes must be the same for each category.)

§12.2 One-Way ANOVA Test

Definition (One-Way ANOVA Test)

A one-way ANOVA test is a method of testing the equality of three or more population means by analyzing sample variances. One-way analysis of variance is used with data categorized with one **factor**, so there is one characteristic used to separate the sample data into the different categories.

- Populations must be approximately normally distributed.
- They must have the same variance, roughly. (Or at least the sample sizes must be the same for each category.)
- The samples must be simple random and quantitative.

§12.2 One-Way ANOVA Test

Definition (One-Way ANOVA Test)

A one-way ANOVA test is a method of testing the equality of three or more population means by analyzing sample variances. One-way analysis of variance is used with data categorized with one **factor**, so there is one characteristic used to separate the sample data into the different categories.

- Populations must be approximately normally distributed.
- They must have the same variance, roughly. (Or at least the sample sizes must be the same for each category.)
- The samples must be simple random and quantitative.
- The samples must be independent - not matched or paired.

§12.2 One-Way ANOVA Test

Definition (One-Way ANOVA Test)

A one-way ANOVA test is a method of testing the equality of three or more population means by analyzing sample variances. One-way analysis of variance is used with data categorized with one **factor**, so there is one characteristic used to separate the sample data into the different categories.

- Populations must be approximately normally distributed.
- They must have the same variance, roughly. (Or at least the sample sizes must be the same for each category.)
- The samples must be simple random and quantitative.
- The samples must be independent - not matched or paired.
- The different samples are from populations that are categorized in only one way.

§12.2 One-Way ANOVA Test

In a one-way ANOVA test, our null hypothesis is always that all the population means are equal. The alternative hypothesis, then, is that at least one is different from the others.

§12.2 One-Way ANOVA Test

In a one-way ANOVA test, our null hypothesis is always that all the population means are equal. The alternative hypothesis, then, is that at least one is different from the others.

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

§12.2 One-Way ANOVA Test

In a one-way ANOVA test, our null hypothesis is always that all the population means are equal. The alternative hypothesis, then, is that at least one is different from the others.

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

H_1 : At least one of the population means differs from the others.

§12.2 One-Way ANOVA Test

In a one-way ANOVA test, our null hypothesis is always that all the population means are equal. The alternative hypothesis, then, is that at least one is different from the others.

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

H_1 : At least one of the population means differs from the others.

You can find "by hand" computations for a one-way ANOVA test, see page 605. The test is a right-tailed F test.

§12.2 Example

The weights (in kg) of oak trees is given below for trees planted in the same plot but with different growth-enhancing methods applied. Use a 0.05 significance level to test the claim that the four treatment categories yield oak trees with the same mean weight. Does there appear to be a best method for enhancing growth of trees in this soil?

Control	Fert	Irr	Fert & Irr
0.24	0.92	0.96	1.07
1.69	0.07	1.43	1.63
1.23	0.56	1.26	1.39
0.99	1.74	1.57	0.49
1.80	1.13	0.72	0.95

§12.2 Example

Control	Fert	Irr	Fert & Irr
0.24	0.92	0.96	1.07
1.69	0.07	1.43	1.63
1.23	0.56	1.26	1.39
0.99	1.74	1.57	0.49
1.80	1.13	0.72	0.95

We use StatCrunch to conduct the test.

§12.2 Example

What's our null and alternative hypotheses? Based on the output, do we reject or accept the null hypothesis? What does this mean? (Use $\alpha = 0.05$.)

Options

Analysis of Variance results:
Data stored in separate columns.

Column statistics

Column	n	Mean	Std. Dev.	Std. Error
Control	5	1.19	0.62573956	0.27983924
Fert	5	0.884	0.624924	0.27947451
Irr	5	1.188	0.34665545	0.15502903
F & I	5	1.106	0.43598165	0.19497692

ANOVA table

Source	DF	SS	MS	F-Stat	P-value
Columns	3	0.3114	0.1038	0.38010491	0.7687
Error	16	4.36932	0.2730825		
Total	19	4.68072			

§12.2 Example

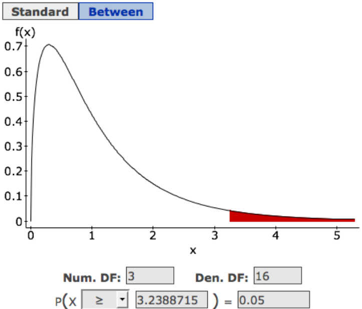
You can also perform the previous test using a critical value test. The F critical value is computed using the significance, the F calculator, and the "degrees of freedom of the numerator" and the "degrees of freedom of the denominator" which are, respectively, one fewer than the number of columns and the difference between the total number of data points and the number of columns.

§12.2 Example

You can also perform the previous test using a critical value test. The F critical value is computed using the significance, the F calculator, and the "degrees of freedom of the numerator" and the "degrees of freedom of the denominator" which are, respectively, one fewer than the number of columns and the difference between the total number of data points and the number of columns. Recall the F stat from the output was 0.3801.

§12.2 Example

You can also perform the previous test using a critical value test. The F critical value is computed using the significance, the F calculator, and the "degrees of freedom of the numerator" and the "degrees of freedom of the denominator" which are, respectively, one fewer than the number of columns and the difference between the total number of data points and the number of columns. Recall the F stat from the output was 0.3801.



§12.3 Two-Way ANOVA Test

Definition (Two-Way ANOVA Test)

A two-way ANOVA test is a method of testing the effect of interactions of two factors and of testing (if the two factors do not interact) the effect of the factors separately.

§12.3 Two-Way ANOVA Test

Definition (Two-Way ANOVA Test)

A two-way ANOVA test is a method of testing the effect of interactions of two factors and of testing (if the two factors do not interact) the effect of the factors separately.

- Populations must be approximately normally distributed.

§12.3 Two-Way ANOVA Test

Definition (Two-Way ANOVA Test)

A two-way ANOVA test is a method of testing the effect of interactions of two factors and of testing (if the two factors do not interact) the effect of the factors separately.

- Populations must be approximately normally distributed.
- They must have the same variance, roughly.

§12.3 Two-Way ANOVA Test

Definition (Two-Way ANOVA Test)

A two-way ANOVA test is a method of testing the effect of interactions of two factors and of testing (if the two factors do not interact) the effect of the factors separately.

- Populations must be approximately normally distributed.
- They must have the same variance, roughly.
- The samples must be simple random and quantitative.

§12.3 Two-Way ANOVA Test

Definition (Two-Way ANOVA Test)

A two-way ANOVA test is a method of testing the effect of interactions of two factors and of testing (if the two factors do not interact) the effect of the factors separately.

- Populations must be approximately normally distributed.
- They must have the same variance, roughly.
- The samples must be simple random and quantitative.
- The samples must be independent - not matched or paired.

§12.3 Two-Way ANOVA Test

Definition (Two-Way ANOVA Test)

A two-way ANOVA test is a method of testing the effect of interactions of two factors and of testing (if the two factors do not interact) the effect of the factors separately.

- Populations must be approximately normally distributed.
- They must have the same variance, roughly.
- The samples must be simple random and quantitative.
- The samples must be independent - not matched or paired.
- The different samples are from populations that are categorized in two ways.

§12.3 Two-Way ANOVA Test

Definition (Two-Way ANOVA Test)

A two-way ANOVA test is a method of testing the effect of interactions of two factors and of testing (if the two factors do not interact) the effect of the factors separately.

- Populations must be approximately normally distributed.
- They must have the same variance, roughly.
- The samples must be simple random and quantitative.
- The samples must be independent - not matched or paired.
- The different samples are from populations that are categorized in two ways.
- All cells have the same number of values.

The null hypothesis is always that there is **no effect due to interaction between the two factors**. The alternative is that there is an effect from the interaction between the two factors. The test statistic is always

$$F = \frac{MS(\text{interaction})}{MS(\text{error})}.$$

Note: If you conclude that the there is an effect due to interaction, then you would not investigate the rows and columns separately.

§12.3 Example

Below are pulse rates of men and women over and under the age of 30. Conduct a two-way ANOVA and state the results. Use a 0.05 significance level.

	< 30	≥ 30
F	78 104 78 64 60 98 82 98 90 96	76 76 72 66 72 78 62 72 74 56
M	60 80 56 68 68 74 74 68 62 56	46 70 62 66 90 80 60 58 64 60

§12.3 Example

We get the following result from StatCrunch.

Analysis of Variance results:
Responses: Results
Row factor: Row Factor
Column factor: Column Factor

ANOVA table

Source	DF	SS	MS	F-Stat	P-value
Row Factor	1	1322.5	1322.5	10.862919	0.0022
Column Factor	1	592.9	592.9	4.8700374	0.0338
Interaction	1	448.9	448.9	3.6872319	0.0628
Error	36	4382.8	121.74444		
Total	39	6747.1			

§12.3 Example

So there does not appear that there is an effect due to the interaction of age and gender, but there does seem to be an effect due to age and a separate effect due to gender.

§12.3 Example

We would like to perform a two-way ANOVA (with $\alpha = 0.05$ throughout) to test whether there is an effect due to interaction and due to the row and column factors separately.

§12.3 Example

We would like to perform a two-way ANOVA (with $\alpha = 0.05$ throughout) to test whether there is an effect due to interaction and due to the row and column factors separately.

	< 30	30 to 50	> 50
G	575 600 600 620	600 620 620 645	620 645 650 650
NG	620 645 650 650	600 620 620 645	575 600 600 620

§12.3 Example

We would like to perform a two-way ANOVA (with $\alpha = 0.05$ throughout) to test whether there is an effect due to interaction and due to the row and column factors separately.

	< 30	30 to 50	> 50
G	575 600 600 620	600 620 620 645	620 645 650 650
NG	620 645 650 650	600 620 620 645	575 600 600 620

Well, there is certainly an effect due to interaction! So we don't perform the hypothesis tests for effects due to the row and column factors separately.

§12.3 Example

Say you conduct a one-way ANOVA on four sample means $\bar{x}_1 = 1.5$, $\bar{x}_2 = 1.6$, $\bar{x}_3 = 1.4$, and $\bar{x}_4 = 1.7$ and got that the P-value was 0.12. What would you conclude?

§12.3 Example

Say you conduct a one-way ANOVA on four sample means $\bar{x}_1 = 1.5$, $\bar{x}_2 = 1.6$, $\bar{x}_3 = 1.4$, and $\bar{x}_4 = 1.7$ and got that the P-value was 0.12. What would you conclude? Would you be able to say which population had the different mean?

§12.3 Example

Say you conduct a one-way ANOVA on four sample means $\bar{x}_1 = 1.5$, $\bar{x}_2 = 1.6$, $\bar{x}_3 = 1.4$, and $\bar{x}_4 = 1.7$ and got that the P-value was 0.12. What would you conclude? Would you be able to say which population had the different mean? Which population would you suspect had the largest mean?

§12.3 Example

Say you conduct a one-way ANOVA on four sample means $\bar{x}_1 = 1.5$, $\bar{x}_2 = 1.6$, $\bar{x}_3 = 1.4$, and $\bar{x}_4 = 1.7$ and got that the P-value was 0.12. What would you conclude? Would you be able to say which population had the different mean? Which population would you suspect had the largest mean? How would you test to see that this population actually had the largest mean?