Learning Goal:
Conduct a hypothesis test for differences in three or more population means using an ANOVA F-test.

Introduction

In this Unit we learn the ANOVA One Way F-test. It allows us to compare means from three or more populations. Alternatively, we can think of this test as examining the relationship between two variables. The explanatory variable is categorical and has 3 or more values. The response variable is quantitative.

For example, we could compare mean sleep hours for four populations: college freshmen, sophomores, juniors, and seniors. Here the explanatory variable is College Class; it has four values (freshmen, sophomore, etc.). The response variable is quantitative: sleep hours.

1) Which of the following scenarios are a candidate for use of the ANOVA?

   - We compare student loan debt for male and female college students.
   - We compare the proportion of college students receiving student loans based on their employment status: not employed, employed part-time, employed full-time.
   - We compare student loan debt for college students based on their academic standing: satisfactory academic progress, academic warning, suspension, reinstatement.

Stating hypotheses

As before, the null hypothesis is a statement of "no difference" or “no relationship between explanatory and response variables” or "no treatment effect." (These all mean the same thing.). We can write the null hypothesis using symbols:

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

Unlike previous tests, where we had three options for the alternative hypothesis (less than, greater than or not equal to), in the ANOVA F-test the alternative hypothesis cannot specify the way in which the means differ. The alternative hypothesis just says, "not all of the population means are equal" or equivalently “at least one population mean differs from the others” or “there is no relationship between the explanatory and response variable” or in the case of an experiment "no treatment effect." (These all mean that same thing.)

2) It is not possible to write the ANOVA’s alternative hypothesis concisely with symbols. Why not?

The F-statistic

We need a way to judge the variation in the sample means to determine if the differences are statistically significant.

- If we find that the sample means are not close together (or at least one deviates substantially from the others), we'll say that we have evidence against H_o and the population means are not equal.
- Otherwise, if the sample means are close together, we'll say that we do not have evidence against H_o , i.e. the differences in the sample means are not statistically significant and the sample means could come from populations with equal means.

The F-statistic measures the variation in sample means relative to the variation in the data within each sample. The F-statistic is very complicated and we will not compute it by hand, but, in essence, it is a ratio of variations (hence the name Analysis of Variance – ANOVA)

$$F = \frac{variation\ in\ sample\ means}{variation\ in\ each\ sample}$$

3) Do elementary school children carry too much weight in their backpacks?

   The explanatory variable is Grade (3^{rd}, 5^{th} or 7^{th}) and the response variable is percent of body weight carried in a school backpack. For example, if a child weighs 40 pounds and carries a 10-pound backpack, PercentWt = 25 because 10 ÷ 40 = 0.25.

   We use ANOVA to test the hypotheses:

   H_0: Mean percent of bodyweight carried in a school backpack is the same for the populations of students in the three grades. (3^{rd}, 5^{th} or 7^{th})

   H_a: At least one population has a mean percent that differs from the others.

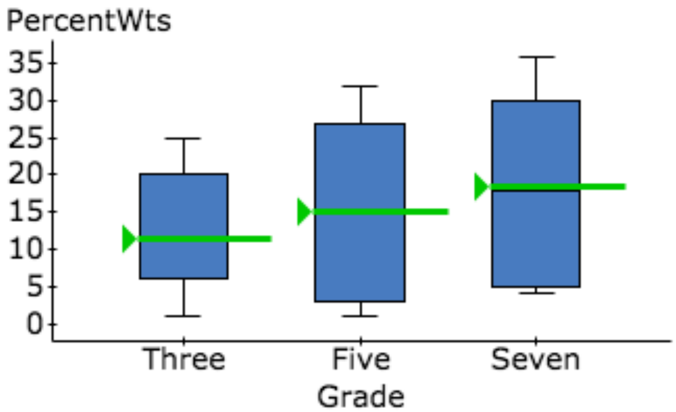
   a) Imagine that we run the ANOVA test as part of two different studies using data from random samples of 3^{rd}, 5^{th} and 7^{th} graders. Examine the boxplots and summary statistics for the hypothetical data from the two studies on the next page.

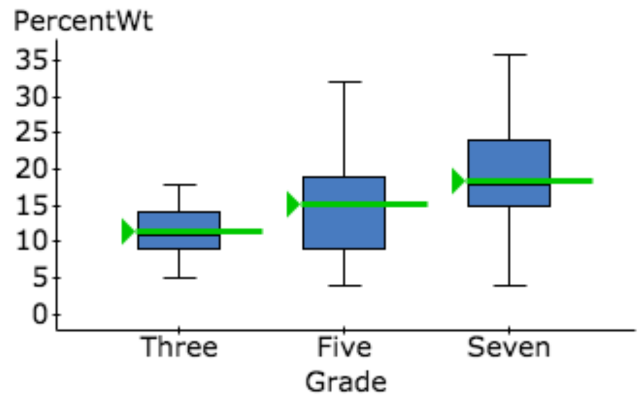
      For one of these studies F = 6.1 with a P-value of 0.0038; for the other study F = 2.3 with a P-value of 0.1098.

      Which is which? Why do you think so?

   b) Which study suggests that the samples come from populations with different means? Why do you think so?

STUDY #1 ***STUDY #2***





Sample means are marked in each boxplot.

Grade	n	Mean	Std. Dev.
Three	21	11.4	7.7
Five	22	15.1	12.4
Seven	19	18.5	10.4

Grade	n	Mean	Std. Dev.
Three	21	11.4	3.7
Five	22	15.1	7.1
Seven	19	18.5	7.8

4) This ratio of variations is the idea behind the comparing more than two means, hence the name Analysis of Variance (ANOVA).

- When the variation within the sample data is large relative to the difference in the sample means, like in (*circle one: Study #1 or Study #2*), the data provide very little evidence against H_0 . In this case, the F-statistic is small and, as we have seen before, a small test statistic has a *(circle one: large or small)* P-value. We (*circle one: fail to reject the null hypothesis or reject the null hypothesis.*) There is *(circle one: weak or strong)* evidence that the population means differ.

- When the variation within the sample data is small relative to the differences in the sample means like in (*circle one: Study #1 or Study #2*), the data give stronger evidence against H_0. In this case, the F-statistic is large and, as we have seen before, a large test statistic has a *(circle one: large or small)* P-value. We (*circle one: fail to reject the null hypothesis or reject the null hypothesis.*) There is *(circle one: weak or strong)* evidence that the population means differ.

Conditions for use of the ANOVA F-test

- Samples are randomly selected from each population and are independent.

- The response variable varies normally within each of the populations. If we cannot verify the shape of the distribution of the response variable in the population, then we can still use the ANOVA in the following situations:

    - If the sample sizes are large ($n > 30$ for each sample), use the ANOVA regardless of the shape of the variable's distribution in the populations.
    - If a sample is small ($n < 30$), examine the distribution of the variable in that sample. If the distribution of the data in each small sample is not heavily skewed and without outliers, this suggests that the variable may be normally distributed in the associated population and we use the ANOVA test.

- This is the new condition: the populations all have the same standard deviation.

    - We will often have difficulty verifying that the population standard deviations are the same; therefore, we check this condition by examining the *sample* standard deviations to see if they are approximately equal. A common rule of thumb is the ratio between the largest sample standard deviation and the smallest is less than 2. If that's the case, we check off this condition as satisfied.

5) Are conditions met for use of the ANOVA F-test in our two fictitious studies of children's backpack weights? Explain.

6) When ANOVA F-test suggests that the population means differ, we can examine confidence intervals estimating each population mean to try to determine which population means account for the difference. These are the same one-sample T-intervals we learned in Unit 9.

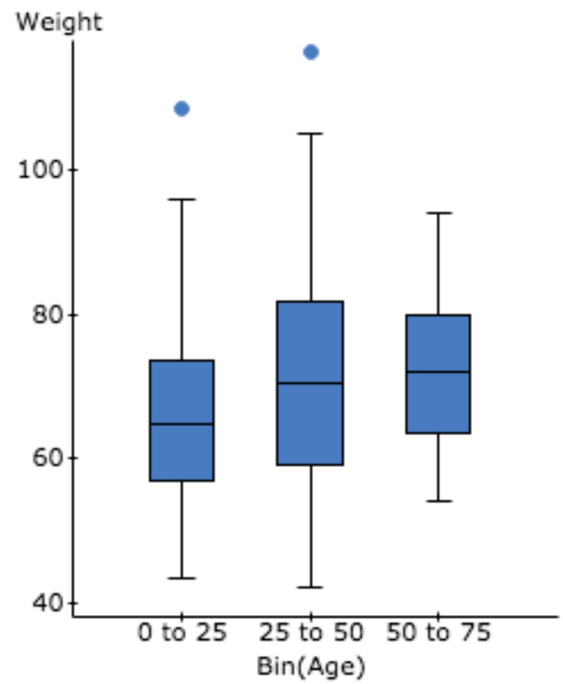
   The fictitious data from Study #2 give these 95% confidence intervals. Which population means appear to differ? Which might be the same?

Grade	Sample Mean	Std. Err.	DF	L. Limit	U. Limit
Three	11.4	0.8	20	9.7	13.1
Five	15.1	1.5	21	11.9	18.3
Seven	18.5	1.8	18	14.8	22.3

(Note: Statisticians actually use more sophisticated statistical techniques called "multiple comparisons" to identify differences, but we will not study those techniques in this introductory statistics course.)

Group Work

1) In a nationwide study of the weights of adults with gym memberships, researchers obtain a random sample of gym members and group them into three age categories. They conduct an ANOVA test and provide the StatCrunch outputs shown here.



Analysis of Variance results:

Responses: Weight (kg)

Factors: Bin(Age)

Response statistics by factor

Bin(Age)	n	Mean	Std. Dev.	Std. Error
0 to 25	180	66.060556	12.163718	0.90663
25 to 50	304	70.750329	13.880549	0.79610408
50 to 75	23	72.121739	10.78846	2.2495494

ANOVA table

Source	DF	SS	MS	F-Stat	P-value
Bin(Age)	2	2699.7154	1349.8577	7.7819725	0.0005
Error	504	87423.629	173.45958		
Total	506	90123.344			

Which conclusions are appropriate at the 5% level of significance? Check all that apply.

- ☐ We cannot draw a conclusion because conditions for the ANOVA F-test are not met.
- ☐ The differences observed in sample mean weights do not provide strong evidence of a difference in population mean weights for the adult gym members in these three age groups.
- ☐ There are statistically significant differences in sample mean weights for the three age groups.
- ☐ This study provides strong evidence that the mean weights differ for the three populations of gym members defined by these age groups.
- ☐ This study provides strong evidence that the mean weight of all gym members who are under 25 is less than the mean weight for all members who are in the 50-75 age group.
- ☐ This study suggests that the mean weights are the same for the three populations of gym members defined by these three age categories.

2) A Statistics student designs a survey for her project. Here are her survey questions: (You do not need to answer the survey questions.)

   How do you get to LMC? (car, bike, bus, walk, other)
   What is your gender? (male, female)
   What is your employment status? (unemployed, employed part-time, employed full-time)
   How many units are you taking this semester?
   What was your high school GPA?
   What is your college GPA?
   How many hours of sleep do you get in a typical night?
   How old are you? (under 20, 20 up to 25, 25 up to 30, 30+)
   Do you have children? (yes, no)
   How many hours do you exercise in a typical week?

   a) From the list of survey questions identify an explanatory variable and a response variable that could be used for an ANOVA test.

   b) State your research question and the hypotheses you would test in the context of the variables you chose.

3) Conduct an ANOVA test using StatCrunch and data from StatCrunchU.

   Use the ANOVA F-test to compare freshmen, sophomores, juniors and seniors (Class) for a quantitative response variable of your choice. You can choose between credit hours, work hours, student loan debt or credit card debt.

   State the hypotheses, check conditions, give summary statistics (means, SDs, n's), give F and the P-value, state your conclusion in context.

   <u>Instructions for accessing the data</u>: Log into StatCrunch. Under **Resources**, under **Take a sample from StatCrunchU**, click on **StatCrunchU**. Scroll down underneath the survey to **Sample size** and select 200. Click Survey.

   You will see a spreadsheet with the survey responses for your random sample of 200 students. To understand the variable defining each column, refer to the survey questions above the spreadsheet.

   StatCrunch Instructions for the ANOVA F-test

   - With the data spreadsheet open, select **Stat**, **ANOVA**, **One Way**
   - Under **Compare,** select **Values in a single column.** For **Responses in** choose the column for response variable you chose. For **Factors in** choose the column with the explanatory variable (Class).
   - Press **Compute!**