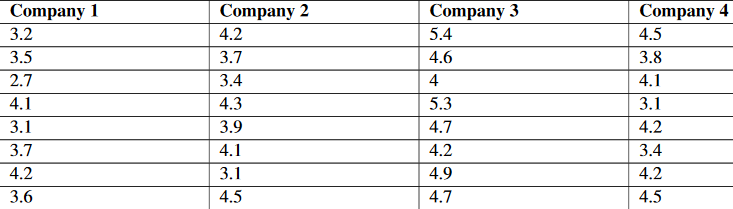
**Analysis of Variance**

* Previously, we’ve discussed analyses that allow us to test if the means + variances of 2 populations are equal.
* Suppose a salesperson wants to compare customer satisfaction for 4 difference insurance companies 🡪 "*Is there a difference in satisfaction scores across the 4 insurance companies*?”
* The satisfaction scores for a sample of customers for each insurance company are recorded:



* We could conduct a series of t-tests to determine if any of the sample means differ.
* However, this would be tedious + has a major flaw, so instead, we use the **Analysis of Variance**
* **(ANOVA)**, which allows us to test the hypothesis that multiple population means + variances of scores are equal.
* The Null and Alternative hypotheses for a one-way ANOVA can be written as:
* H(0): Means of all factor levels are equal
* H(A): At least 1 factor level has a different mean
* ANOVA can be used when we want to test the means of 3+ populations at once.
* Theoretically, we could test hundreds of population means using this procedure.
* This ANOVA is technically called “one-way” as it has just 1 main grouping factor: *company*.

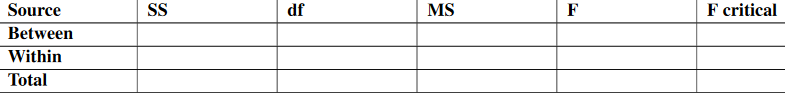
Shortcomings of Comparing Multiple Means Using Previously Explained Methods

* Why should we learn a new test called ANOVA, when we could just conduct a series of t-tests?
* To answer our question, we could just run 6 different independent samples t-tests
* We could use a CPU to compute these quickly and easily.
* *It turns out this is a very bad idea, and has a major flaw:*
* When more than 1 t-test is run, each at its *own* level of significance, the probability of making 1 or more Type I errors multiplies *exponentially* (reject the null when we should not)
* The level of significance, α, is the probability of a Type I error in a *SINGLE test*.
* So, for a single t-test w/ α = 0.05, we have a Type I error probability of 5%.
* When testing *more* than 1 pair of samples, the probability of making at least one Type I error is:



Where c is the number of independent t-tests.

* If our salesperson conducted the 6 separate t-tests to examine the means of the populations w/ α = 0.05, the probability of committing a Type I error would be 0.265 or 26.5%.
* Assumptions of the ANOVA test
* All observations are independent of one another + randomly selected from the population which they represent.
* The population at each factor level is approximately normal.
* The variances for each factor level are approximately equal to one another.
* W/ ANOVA, we’re actually analyzing the TOTAL variation of the scores, including the variation of the scores WITHIN the groups and the variation BETWEEN the group means.
* Since we’re interested in 2 different types of variation, we 1st calculate each type of variation independently + then calculate the ratio between them, an **F-value**.
* Just like z-score, t-test, + chi-square tests, ANOVA has its own distribution to use to set our critical values + test our hypothesis.
* Just like the t + chi-square distributions which use dF, the F-distribution also relies on dF.
* Since the F-value is actually a *ratio* of 2 different sources of variance, we’ll need 2 *different* dF
* *When using ANOVA, we are testing null that the means + variances of our samples are equal.*
* When conducting a hypothesis test, we’re testing the probability of obtaining an extreme F-statistic by chance.
* If we reject the null that the means variances of the samples are equal, then we’re saying the difference we see could not have happened just by chance.
* To test a hypothesis using ANOVA, there are several steps that we need to take + we can employ a nice little tool called the ANOVA table:



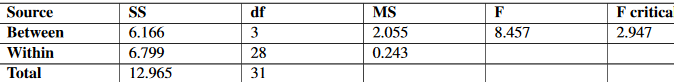
* **1. Calculate the total sum of squares (SS(t)).**
* This is the difference between each score and the **grand mean**.



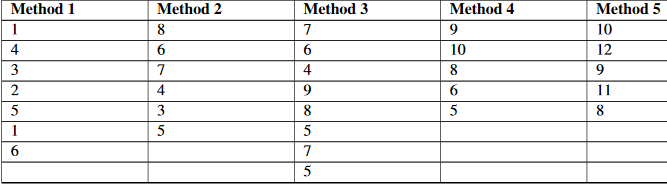
* where: y = each observation, N = total number of scores, y^2 = grand mean
* **2. Calculate the sum of squares between (SS(b)).**



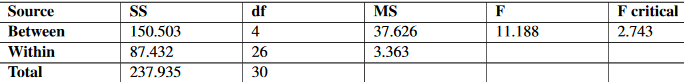
* where k = the # of groups, n(k) = the # of scores in group k, y(k) = the mean of group k
* **3. Find the sum of squares within groups (SS(w))**
* by subtracting SS(w) = SS(t) – SS(b)
* **4. Next solve for degrees of freedom for the test:**
* dF(t) = N – 1 dF(b) = k – 1 dF(w) = N - k
* **5. Using the values, you can now calculate the Mean Squares Between (MS(b) + Mean Squares Within (MS(w)) using the relationships below:**
*  
* **6. Finally, calculate the F statistic using the following ratio:**
* F = MS(b) / MS(w)
* **7. Find F critical**
* For our example, we have 3, 28 dF, so our F-critical value is 2.947
* **8. Fill in the Table from here**



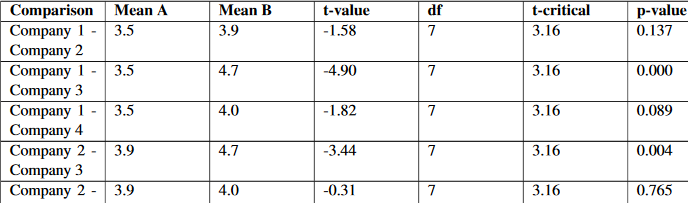
* **9. Interpret the results of the hypothesis test**
* Our F-value from the ANOVA test is greater than the F-critical value, so we would reject our Null + can conclude that the average customer satisfaction scores of the 4 insurance companies are NOT equal to one another – at least ONE is different from the others.
* Ex: Teacher testing multiple reading programs to determine the impact on student achievement. There are 5 different reading programs + 31 students are randomly assigned to 1 of the 5 programs.



* **Step 1: Clearly state the Null and Alternative Hypotheses.**
* H(0): Mean student achievement is the same across all 5 reading programs
* H(A): At least 1 reading program has a mean student achievement level different from the others.
* **Step 2: ID the appropriate significance level and confirm test assumptions.**
* We’ll choose the default significant level of 0.05.
* 1. All observations ARE independent of 1 another + randomly selected from the population which they represent.
* 2. The population at each factor level is approximately normal.
* We assume student achievement level (in the population) is normally distributed. We can check boxplots here to confirm none of the samples are highly skewed.
* 3. The variances for each factor level are approximately equal to one another.
* We check the variance of each group, + they range from 2.5 to 4.3. Since 4.3 < 2\*2.5
* *We can proceed.*
* **Step 3: Analyze the data.**
* We’ll use the 1-way ANOVA test b/c we want to compare 5 different independent methods of reading programs.
* We can find the critical value of F with (4,26) dF to be 2.743.



* **Step 4: Interpret your results.**
* Since our calculated F-value is greater than the F-critical, we can reject our Null + conclude that all 5 reading program means are NOT equal to one another, but that there is at least ONE reading method score mean that is not like the others.
* Now that we’ve found a way to test the Null Hypothesis when we want to compare 3+ population means, we need to learn one more step. *What happens when we reject the Null Hypothesis?*
* If we have an ANOVA that rejects the Null, we must find out WHERE the difference lies/what or which group(s) are different from one another.
* To do this we use a **post-hoc test 🡪** literally run another analysis after the ANOVA shows a rejection of the Null
* The easiest post-hoc analysis (there are several) to run is the **Bonferroni post-hoc analysis**, which is really just all possible t-tests to compare the groups.
* Now, you might want to bring up the problem of inflating the Type I error w/ all those t-tests, but that’s where Bonferroni comes in
* *It’s simply a correction to the level of significance, α, when we run these post-hoc tests.*
* Ex: For the Insurance Company example, we rejected the Null, so we are allowed to run the post-hoc tests to discover more about the difference in the group means.
* We have 4 groups, so we’ll need to run: **(4 \* (4-1)) / 2) = 6** group comparison t-tests.
* The Bonferroni correction *simply divides the significance level used for the ANOVA = 0.05 by 6*.
* So our new significance level (just for the post-hoc comparisons) = 0.008.
* If we were to do these post-hoc comparisons *by hand*, we’d need to run 6 independent samples t-tests + find the critical value for each comparison, based on the dF for the comparison + the new significance level of 0.008.
* Most of the time, these comparisons are done using technology, so we can take advantage of the p-value reported for each t-test.
* Using the p-value reported, we’d compare *that value* to the *new* Bonferroni corrected significance level.
* If the reported p-value of the comparison was < the corrected significance level, we reject the Null that the comparison means are equal.
* Here are the results of each comparison, along with w/ mean of the groups, the t-value, the comparison df, t-critical value, and the p-value:





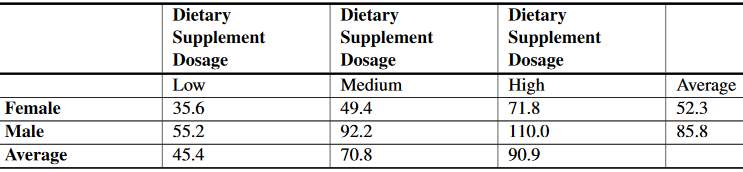
* Our post-hoc comparisons show us that Company 3 was significantly higher in customer satisfaction than Company 1 + Company 2, but NOT Company 4.
* Also, no other Company was *significantly* different from the others.

Summary

* When testing multiple independent samples to determine if they come from the same population, we could conduct a series of separate t-tests in order to compare all possible pairs of means.
* However, a more precise and accurate analysis is the **Analysis of Variance (ANOVA).**
* In ANOVA, we analyze the total variation of the scores, including the **variation of scores within the groups**, **variation between group means**, + the **total mean** of all the groups (the grand mean).
* In this analysis, we calculate the **F-value**, which is the ratio of **mean of squares** **between groups** divided by the **mean of squares within groups**.
* If we are able to reject our Null, we continue on, conducting **post-hoc analyses** to discover where the difference in the sample means lies.

***The Two-Way ANOVA Test***

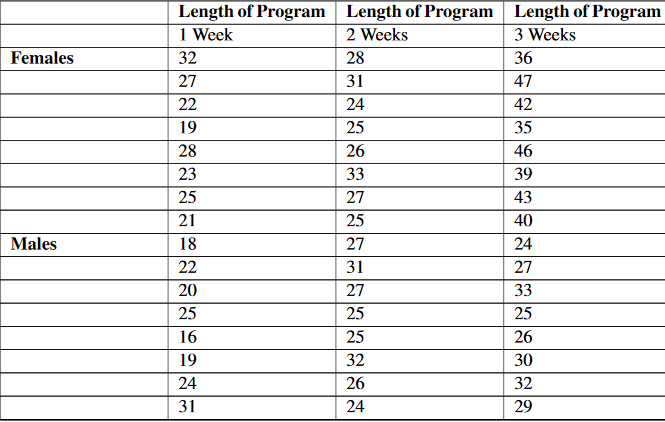
* The one-way ANOVA method is the procedure for testing the null that the population means + variances of a single independent variable are equal.
* Sometimes, however, we are interested in testing the means and variances of *more than one independent variable*.
* Say, for example, a researcher is interested in determining the effects of different dosages of a dietary supplement on the performance of both males + females on a physical endurance test.
* The 3 different dosages of the medicine are low, medium, + high, w/ genders male + female.



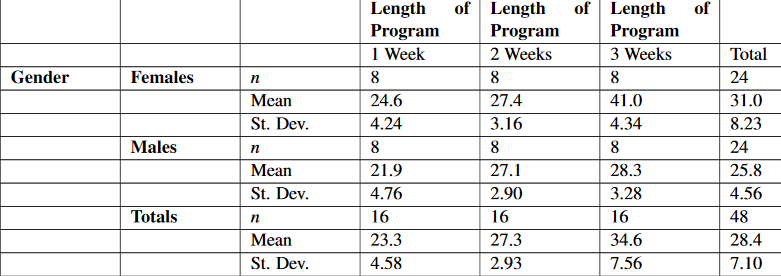
* There are several questions that can be answered by a study like this, such as, "Does the medication improve physical endurance, as measured by the test?" and "Do males females respond in the same way to the medication?"
* While there are similar steps in performing 1-way and 2-way ANOVA tests, there are also some major differences.
* 1-way ANOVA allows us to examine the effect of a single independent variable on a dependent variable (i.e., effectiveness of a reading program on student achievement).
* W/ 2-way ANOVA, we’re not only able to study the effect of 2 independent variables (i.e., effect of dosages and gender on the results of a physical endurance test), but also the interaction between these variables.
* An example of interaction *between* the 2 variables gender and medication is a finding that *men and women respond differently* to the medication.
* We could conduct 2 separate 1-way ANOVA tests to study the effect of the 2 independent variables, but there are several advantages to conducting a two-way ANOVA test.
* Efficiency 🡪 W/ simultaneous analysis of 2 independent variables, a 2-way ANOVA is really carrying out 2 separate research studies at once.
* Control 🡪 When including an *additional* independent variable in the study, we are able to control for that variable.
* Say we included IQ in the example about the effects of a reading program on student achievement.
* By including this variable, we are able to determine the effects of various reading programs, the effects of IQ, *and the possible interaction between the two*.
* Interaction 🡪 W/ a 2-way ANOVA, its possible to investigate the interaction of 2+ independent variables.
* In most real-life scenarios, variables DO interact w/ one another.
* Therefore, the study of the interaction between independent variables may be just as important as studying the interaction between the independent and dependent variables.
* When we perform 2 separate 1-way ANOVA tests, we run the risk of losing these advantages.

Two-Way ANOVA Procedures

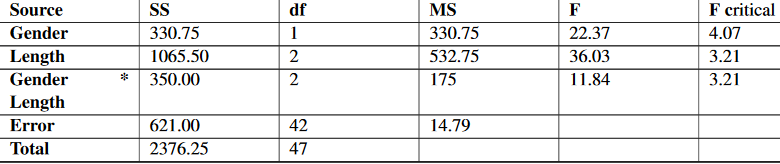
* There are 2 kinds of variables in all ANOVA procedures: dependent and independent.
* In 1-way ANOVA, we were working w/ 1 independent variable + 1 dependent variable.
* In 2-way ANOVA, there are 2 independent variables + 1 dependent variable.
* *Changes in the dependent variables are assumed to be the result of changes in the independent variables.*
* In 1-way ANOVA, we calculated a ratio that measured the variation between the 2 variables (dependent + independent).
* In 2-way ANOVA, we need to calculate a ratio that measures not only the variation *between the dependent and independent variables*, but also the interaction *between the 2 independent variables*.
* Before, when we performed the 1-way ANOVA, we calculated the total variation by determining the variation *within groups* and the variation *between* *groups*.
* Calculating the total variation in two-way ANOVA is similar, but since we have an additional variable, we need to calculate 2 more types of variation.
* Determining the total variation in 2-way ANOVA includes calculating
* Variation *within the group* (within-cell variation)
* Variation in the *dependent variable attributed to 1 independent variable* (variation among row means)
* Variation in the *dependent variable attributed to the other independent variable* (variation among column means)
* Variation *between the independent variables* (the **interaction effect**).
* The formulas we use to calculate these types of variations are very similar to the ones used in the 1-way ANOVA.
* For each type of variation, we want to calculate the **total sum of squared deviations** (**sum of squares**) around the **grand mean**.
* After we find the SS(t), we want to divide it by the dF to arrive at the **mean of squares**, which allows us to calculate our final ratio.
* We could do these calculations by hand, but we have technological tools, such as CPU programs that can compute these figures much more quickly and accurately than we could manually.
* The process for determining + evaluating the null hypothesis for the 2-way ANOVA is very similar to the same process for the 1-way ANOVA.
* However, for the 2-way ANOVA, we have *additional hypotheses, due to the additional variables*.
* For 2-way ANOVA, we have 3 null hypotheses:
* In the population, the means for the rows equal each other.
* Ex: Mean for males = mean for females.
* In the population, the means for the columns equal each other.
* Ex: Means for the 3 dosages are equal.
* In the population, the null hypothesis would be that there is no interaction between the 2 variables (all effects equal 0)
* Ex: There is no interaction between gender + amount of dosage
* Ex: Gym teacher is interested in the effects of the length of an exercise program on the flexibility of male + female students.
* The teacher randomly selected 48 students (24 males, 24 females) + assigned them to exercise programs of varying lengths (1, 2, or 3 weeks).
* At the end of the programs, she measured the students’ flexibility recorded the following results.



* *Do gender + the length of an exercise program have an effect on the flexibility of students?*



* It appears that females have more flexibility than males and that the longer programs are associated w/ greater flexibility.
* Also, we can take a look at the standard deviation of each group to get an idea of the variance *within groups*.
* This info is helpful, but it is necessary to calculate the test statistic to more fully understand the effects of the independent variables and the interaction between these 2 variables



* Note that the CPU finds the dF for **interaction** by multiplying together the dF for each variable (rows + columns).
* From this summary table, we can see that all 3 F -ratios exceed their respective critical values.
* *This means that we can reject all three null hypotheses and conclude that:*
* In the population, the *mean for males differs from the mean of females*.
* In the population, the *means for the three exercise programs differ*.
* There IS an *interaction between the length of the exercise program and the student’s gender*.

Experimental Design + its Relation to the ANOVA Methods

* **Experimental design** is the process of taking the time + effort to organize an experiment so that the data are readily available to answer the questions that are of most interest to the researcher.
* When conducting an experiment using ANOVA, there are several ways that we can design an experiment.
* The design we choose depends on the nature of the questions we are exploring.
* In a **totally randomized design**, subjects/objects are assigned to treatment groups completely at random
* Ex: teacher might randomly assign students into 1 of 3 reading programs to examine the effects of the different reading programs on student achievement.
* Often, the person conducting the experiment will use a CPU to randomly assign subjects.
* In a **randomized block design**, subjects/objects are 1st divided into homogeneous categories before being randomly assigned to a treatment group
* Ex: athletic director studying the effect of various physical fitness programs on males + females, he would 1st categorize randomly selected students into homogeneous categories (males + females) before randomly assigning them to 1 of programs he was trying to study.
* In ANOVA, we use *both* randomized design + randomized block design experiments.
* In 1-way ANOVA, we typically use a **completely randomized design**.
* By using this design, we can assume the observed changes are caused by changes in the independent variable.
* In 2-way ANOVA, since we are evaluating the effect of 2 independent variables, we typically use a **randomized block design**.
* Since the subjects are assigned to 1 group + then another, we are able to evaluate the effects of *both* variables and the interaction between the two.

*Lesson Summary*

* With 2-way ANOVA, we are not only able to study the effect of 2 independent variables, but also the interaction between these variables.
* There are several advantages to conducting a two-way ANOVA, including efficiency, control of variables, + the ability to study the interaction between variables.
* Determining the total variation in 2-way ANOVA includes calculating the following:
* Variation within the group (within-cell variation)
* Variation in the dependent variable attributed to one independent variable (variation among the row means)
* Variation in the dependent variable attributed to the other independent variable (variation among the column means)
* Variation between the independent variables (the interaction effect)
* It is easier and more accurate to use technological tools to calculate the figures needed to evaluate our hypotheses tests.