One Factor Analysis of Variance (ANOVA)

In the Section 9.4.1, we used the statistical inference model to compare two population means with independent sampling. This models can be expanded to compare more than two independent populations using a technique called Analysis of Variance, or ANOVA for short. There are many ANOVA models, but we limit our study to one of them, the One Factor ANOVA model, also known as One Way ANOVA.

Party Pizza Example

Party Pizza specializes in meals for students. Hsieh Li, President, recently developed a new vegan tofu pizza. Before making it a part of the regular menu she decides to test it in several of her restaurants. She would like to know if there is a difference in the mean number of tofu pizzas sold per day at the Cupertino, San Jose, and Santa Clara pizzerias. Data was collected for several days at each location .

Location	Cupertino	San Jose	Santa Clara
Tofu Pizzas sold each day	13	10	18
	12	12	16
	14	13	17
	12	11	17
			17

Factorial Design – Indentifying the variables

One way of looking at this model is considering a single population with one **numeric** and one **categorical** variable being sampled. The numeric variable is called the **response** and the categorical variable is the **factor**. The goal in one factor ANOVA is to show that the **response** will depend on the factor.

1. In the Party Pizza example, what is the numeric **response** variable?

2. In the Party Pizza example, what is the categorical **factor** variable?

3. The Hypotheses can be written in context by putting the response in the first blank and the factor in the second blank of each statement.

Ho: There is no difference in mean _____

(response in context)

due to _____.

(factor in context)

Ha: There is a difference in mean _____

(response in context)

due to _____.

(factor in context)

4. The choices for the factor are called the **levels** and the number of levels is denoted by the letter k. For the Party Pizza example, identify the levels and determine the value of k.

5. The number of observations per level are called the replicates (n_1=4, n_2=4, n_3=5 in our example). If the replicates are equal, the design is **balanced.** The sum of the replicates is called the sample size (n = 4 + 4 + 5 = 13). Is the Party Pizza Example balanced or unbalanced?

Hypotheses in terms of population parameters

Suppose we wanted to compare the means of more than two (k) independent populations and want to test the null hypothesis $Ho: \mu_1 = \mu_2 = \cdots = \mu_k$. If we can assume all population variances are equal, we can expand the pooled variance t-test for two populations to one factor ANOVA for k populations. The alternative hypothesis is always written as Ha: At least one of the means is different.

6. Write the hypotheses for the Party Pizza example using population parameters

Ho: _____

Ha: _____

The logic of ANOVA - How comparing variances will test for a difference in means.

It may seem strange to use a test of "variances" to compare means, but this graph demonstrates the logic of the test. If the null hypothesis $Ho: \mu_1 = \mu_2 = \mu_3$ is true, then each population would have the same distribution and the variance of the combined data would be approximately the same.

However, if the Null Hypothesis is false, then the difference between centers would cause the combined data to have an increased variance. We can use a test statistic called the F distribution that uses sample data to test the Null Hypothesis.

Ho is true -all means the same

Ho is false -not all means the same
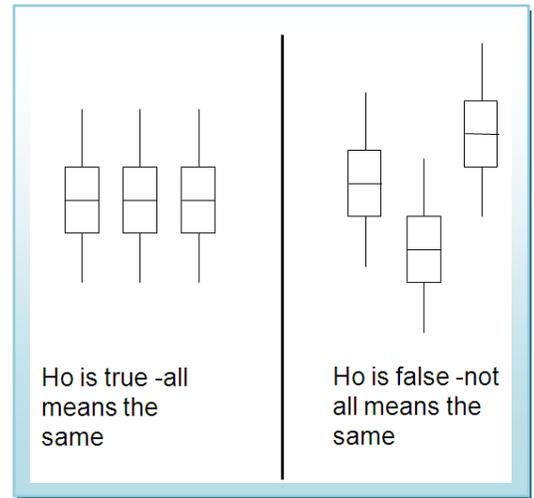
7. From the Party Pizza example, calculate the sample mean for each of the three locations.

Cupertino: $\overline{X}_1 = \dfrac{13+12+14+12}{4} = 12.75$

San Jose: $\overline{X}_2 = \dfrac{}{} =$

Santa Clara: $\overline{X}_3 = \dfrac{}{} =$

8. Now combine the data of all three locations and calculate the Overall Mean.

Overall Mean: $\overline{\overline{X}} = \dfrac{}{13} =$

9. Compare the means from each location to the Overall Mean. Which location deviates the most from the Overall Mean?

The One Factor ANOVA model

Recall the sample variance which measures variability with respect to the mean and think of it as the sum of square deviations from the mean divided by the degrees of freedom.

$$s^2 = \frac{\sum (X_i - \overline{X})^2}{n-1} = \frac{Total\ Sum\ of\ Squared\ Deviations}{Total\ Degrees\ of\ Freedom} = \frac{SS_{total}}{df_{total}}$$

In ANOVA, We can then partition the total sum of squared deviations and total degrees of freedom into two components: The portion of variability **between** the sample means of each group (**SS$_{Factor}$**) and the portion of variability **within** each group (**SS$_{Error}$**). We also partition the degrees of freedom:

$$SS_{Total} = SS_{Factor} + SS_{Error} \qquad df_{Total} = n-1 \quad df_{Factor} = k-1 \quad df_{Error} = n-k$$

We then divide these partitioned sums of squares by the matching degrees of freedom to get two estimates of the variance, called **mean squares (MS):**

$$MS_F = SS_{Factor}/df_{Factor} \qquad\qquad MS_E = SS_{Error}/df_{Error}$$

The mean square factor (MS$_F$), also known as mean square between, measures the variability of the means **between** groups, while the mean square error (MS$_E$), also known as mean square within, measures the variability within each sample. Under the null hypothesis, the ratio of MS$_F$/MS$_E$ should be close to 1 and has **F distribution**. The family of F distributions are skewed positive and defined by two different degrees of freedom, df$_{Num}$ and df $_{Den}$ (numerator and denominator degrees of freedom). One Factor ANOVA is always calculated as a right-tailed test using the F distribution.
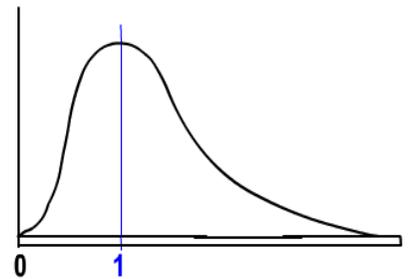
One Factor ANOVA model to compare the means of k independent populations

Model Assumptions	Test Statistic
• The populations being sampled are normally distributed. • The populations have equal standard deviations. • The samples are randomly selected and are independent.	$F = \dfrac{MS_{Factor}}{MS_{Error}}$ $df_{num} = k - 1$ $df_{den} = n - k$

The ANOVA table

When running Analysis of Variance, the data is usually organized into a special ANOVA table, especially when using computer software.

Source of Variation	Sum of Squares (SS)	Degrees of Freedom (df)	Mean Square (MS)	F
Factor (Between)	SS$_{Factor}$	k-1	MS$_F$= SS$_{Factor}$/k-1	F= MS$_F$/MS$_E$
Error (Within)	SS$_{Error}$	n-k	MS$_E$= SS$_{Error}$/n-k	
Total	SS$_{Total}$	n-1		

Sum of Squares: The total variability of the numeric data being compared is broken into the variability between groups (SS_{Factor}) and the variability within groups (SS_{Error}). These formulas are the most tedious part of the calculation. T_c represents the sum of the data in each population and n_c represents the sample size of each population (replicates). These formulas represent the numerator of the variance formula. We will use technology to expedite the tedious calculations of Sums of Squares.

$$SS_{Total} = \Sigma(X^2) - \frac{(\Sigma X)^2}{n} \qquad SS_{Factor} = \Sigma\left(\frac{T_c^2}{n_c}\right) - \frac{(\Sigma X)^2}{n} \qquad SS_{Error} = SS_{Total} - SS_{Factor}$$

Degrees of freedom: The total degrees of freedom is also partitioned into the Factor and Error components.

Mean Square: This represents calculation of the variance by dividing Sum of Squares by the appropriate degrees of freedom.

F: This is the test statistic for ANOVA: the ratio of two sample variances (mean squares) that are both estimating the same population value has an F distribution. Computer software will then calculate the p-value to be used in testing the Null Hypothesis that all populations have the same mean.

The research hypotheses from the Party Pizza example:
Ho: $\mu_1 = \mu_2 = \mu_3$ (There is no difference in mean tofu pizza sales due to location)
Ha: At least one mean is different (There is a difference in mean tofu pizza sales due to location)

10. From the Party Pizza example: $SS_{Total} = 86$ and $SS_{Factor} = 76.25$. Complete the ANOVA table and determine the F test statistic.

Source of Variation	Sum of Squares (SS)	Degrees of Freedom (df)	Mean Square (MS)	F
Factor (Between)				
Error (Within)				
Total				

We will use technology to determine the p-value of getting an F statistic this extreme if Ho is true. The F value of 39.1 you calculated in question 10 has a p-value of 0.001.

11. At the .05 significance level, would you reject or fail to reject the Null Hypothesis?

12. Can Hsieh Li conclude that there is a difference in the mean number of tofu pizzas sold per day at the three pizzerias? Write a conclusion in the context of the problem.

Post-hoc Analysis – Tukey's Honestly Significant Difference (HSD) Test.

When the Null Hypothesis is rejected in one factor ANOVA, the conclusion is that not all means are the same. This however leads to an obvious question: Which particular means are different? Seeking further information after the results of a test is called post-hoc analysis.

The problem of multiple tests

One attempt to answer this question is to conduct multiple pairwise independent same t-tests and determine which ones are significant. We would compare μ_1 to μ_2, μ_1 to μ_3, μ_2 to μ_3, μ_1 to μ_4, etc. There is a major flaw in this methodology in that each test would have a significance level of α, so making Type I error for the combined tests would be significantly more than the desired α. Furthermore, these pairwise tests would NOT be mutually independent. There were several statisticians who designed tests that effectively dealt with this problem of determining an "honest" significance level of a set of tests; we will cover the one developed by John Tukey, the Honestly Significant Difference (HSD) test.

The Tukey HSD test

Tests: $H_o : \mu_i = \mu_j$ $\qquad$ $H_a : \mu_i \neq \mu_j$ where the subscripts i and j represent two different levels of the factor.

Overall significance level of α. This means that **all pairwise tests** can be run at the same time with an overall significance level of α.

Test Statistic: Tukey's HSD

Decision: Reject Ho if $\left| \overline{X}_i - \overline{X}_j \right| > HSD$ critical value

Computer software, such as Minitab, will identify which pairs of sample means are significantly different.

Let us return to the Party Pizza example where we rejected the Null Hypothesis and supported the claim that there was a difference in means among the three restaurants.

In reviewing the graph of the sample means, it appears that Santa Clara has a much higher number of sales than Cupertino and San Jose.

Individual Value Plot of Cupertino, San Jose, Santa Clara
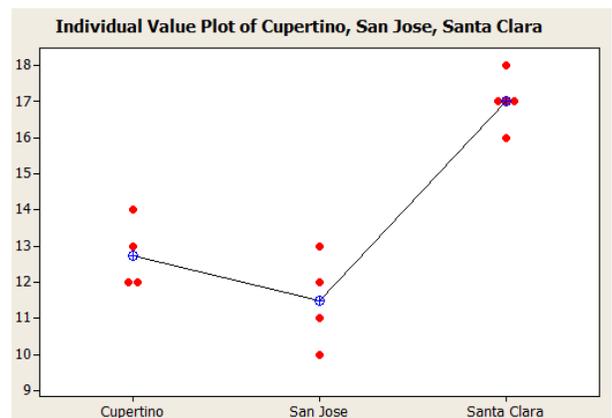
13. There will be three pairwise post-hoc tests to run. For example, we can determine if there is a significant difference in mean pizza sales between the Cupertino and Santa Clara locations.

$$H_o : \mu_1 = \mu_3 \quad H_a : \mu_1 \neq \mu_3$$

Determine Ho and Ha for the other two pairwise comparisons.

These three tests will be conducted with an overall significance level of α = 5%.
The model will be the Tukey HSD test.

The Minitab approach for the decision rule will be to reject Ho for each pair that does not share a common grouping letter.

Example - Data/Results/Conclusion

Refer to the Minitab output. Santa Clara is in group A while Cupertino and San Jose are in Group B.

```
Grouping Information Using Tukey Method

             N     Mean  Grouping
Santa Clara  5  17.0000  A
Cupertino    4  12.7500     B
San Jose     4  11.5000     B

Means that do not share a letter are significantly different.
```

Testing $H_o : \mu_1 = \mu_3 \quad H_a : \mu_1 \neq \mu_3$, we recognize that Cupertino and Santa Clara do not share a common grouping letter. Therefore we reject Ho and conclude that Santa Clara has a higher number of mean tofu pizzas sold then Cupertino.

14. Determine the decision and conclusion for the test comparing Cupertino to San Jose.

15. Determine the decision and conclusion for the test comparing San Jose to Santa Clara.

Are there Differences between Pain Relievers?

An experiment was conducted to compare five treatments for reducing fever. Thirty-five patients were randomly assigned into five treatment groups. The patients' temperatures were taken two hours after receiving the treatment, and are given below. We will test the claim that there is no difference in mean body temperature among each of the five treatment groups using a 5% significance level.

Because of the requirements for this test, we assume that independent samples were drawn from normal populations with equal variances.

Placebo		Aspirin		Anacin		Tylenol		Bufferin
98.3		96.0		97.2		95.1		96.1
98.7		96.6		97.4		95.7		96.5
98.1		95.3		96.9		94.4		95.4
99.9		98.1		97.8		97.2		98.2
96.2		94.6		94.7		93.7		94.5
98.6		96.6		97.0		95.7		96.5
98.3		96.2		96.8		95.3		96.2

1. What is the numeric **response** variable?

2. What is the categorical **factor** variable?

3. Identify the **levels** of the factor and determine the value of **k**.

4. Determine the replicates per level and the total sample size. Is the design balanced or unbalanced?

5. State the Null and Alternative Hypotheses in words.

6. State the Null and Alternative Hypotheses using population parameters.

7. Determine the sample mean for each type of pain reliever. Identify the pain relievers with the highest and lowest means.

8. Complete the Missing Values of the ANOVA table

Source of Variation	Sum of Squares (SS)	Degrees of Freedom (df)	Mean Square (MS)	F	p-value
Factor (Between)	34.63				0.001
Error (Within)	35.63			███	███
Total	70.26		███		

9. At the .05 significance level, would you reject or fail to reject the Null Hypothesis?

10. State your conclusion in the context of the problem.

11. Here is Minitab output showing the results of all 10 pairwise comparisons using a Tukey HSD test. Which **pairs** of pain relievers show a significant difference in mean temperature?

Tukey Pairwise Comparisons

Grouping Information Using the Tukey Method and 95% Confidence

Treatment	N	Mean	Grouping
Placebo	7	98.300	A
Anacin	7	96.829	A B
Bufferin	7	96.200	B
Aspirin	7	96.200	B
Tylenol	7	95.300	B

Means that do not share a letter are significantly different.