C207 Module 3 Notes:

<u>3.03 Linear Programming</u>

- **Linear Programming**: Mathematical technique used for determining how to achieve the best outcome in a given situation.
    - Often used to find the combination of products that will maximize profits or minimize costs.
    - Useful in planning because it allows for a systematic comparison of alternatives to determine which set of choices is best for meeting an objective.
    - Linear programming might be used to schedule the operations of multiple machines or to prioritize expenditures on raw materials.
    - Determining the objective is a key step when applying linear programming to a problem.
    - **CONSTRAINTS**: The state, quality, or sense of being restricted to a given course of action or inaction. An application restriction or limitation, either internal or external to a project, which will affect the performance of the project or a process.
    - **THE SIMPLEX METHOD**: A complicated mathematical method that helps solve linear programming problems.

Application: The Product Mix Problem

- Step 1: Understand the Problem
- Step 2: Calculate Profit per Unit
- Step 3: Identify the Constraints
- Step 4: Solve

<u>3.04 Crossover Analysis</u>

Crossover Analysis

- **CROSSOVER ANALYSIS:** Allows a decision maker to identify the crossover point, which represents the point in which we are indifferent between the plans.
    - Also known as indifference point

<u>3.05 Breakeven Analysis</u>

Break-even Analysis

- **BREAK-EVEN ANALYSIS:** A calculation of the point at which revenues equals expenses.
    - Used to examine various combinations of prices and quantities to determine the point at which revenues cover costs.
    - Two calculations are commonly used in break even analysis.
        - The calculation of break even units, or break even points, is used to determine the volume of unit production necessary to cover production costs.
        - The calculation identifies the volume of sales dollars that must be generated before earning a profit.

- **BREAK-EVEN POINT (BEP):** The volume of sales at which total revenues and total costs are equal.
    - Indicates the point at which the business is no longer operating at a loss. In other words, for every unit sold above break-even, the business is generating positive pretax profits.
- **BREAK-EVEN REVENUE**: The amount of money generated from the sale of the break-even quantity of units.
    - Amount of money generated from the sale of the break-even quantity of units. This number tells a manager how much revenue must be generated before the company will begin to make a profit.

Hypothesis Testing

- **HYPOTHESIS**: A proposed explanation used as a starting point for future examination.
    - Statement or claim about a given population.
- To test a hypothesis, you must convert the question into a NULL HYPOTHESIS and ALTERNATIVE HYPOTHESIS.
    - **NULL HYPOTHESIS:** The argument that there is no difference between two samples or that a sample has not changed over time.
    - For whatever relationship is being tested, the null hypothesis is the statement that the relationship does not exist.
    - For example, if a test is conducted to determine a difference between two means, the null hypothesis will state that there is no difference between these two means.
        - There is NO RELATIONSHIP!
    - **ALTERNATIVE HYPOTHESIS:** The argument that typically states that the sample is not equal to the hypothesized null sample.
        - The alternative hypothesis, or HA, is the opposite statement to the null hypothesis. It states that there is a relationship for whatever relationship is being tested. If we conduct a test to determine the difference between two means, the alternative hypothesis states that there is a difference between these two means.

Statistical Significance

- **STATISTICALLY SIGNIFICANT:** A result that is unlikely to be caused by random variation or errors.
    - A statistically significant result is unlikely to be caused by random variation or errors. A difference that is statistically significant is called a SIGNIFICANT DIFFERENCE. A hypothesis test will tell us whether the results are statistically significant or not.
- **SIGNIFICANT DIFFERENCE:** A difference that is found to be statistically significant. The difference is unlikely to have been caused by random variation or errors.

Writing Null and Alternative Hypotheses

Writing Null and Alternative Hypotheses

Hypothesis testing can be used in many contexts. For example, you can execute a hypothesis test to determine whether two means, denoted μ_1 and μ_2, are *significantly different* from one another.

So, the null hypothesis statement can be written as:

$$H_0: \mu_1 = \mu_2$$

The alternative hypothesis can take the following form:

$$H_A: \mu_1 \neq \mu_2$$

The null hypothesis is the statement that is being tested. There are two possibilities after conducting a hypothesis test:

- Reject the null hypothesis.
- Fail to reject the null hypothesis.

Notice that both of these possibilities pertain to the null hypothesis. The null hypothesis is always the statement that is being tested. The outcome of your experiment is to determine whether the null hypothesis should be rejected. If you reject the null hypothesis, the difference being tested is significant: the difference is most likely not caused by random variation or error. On the other hand, if you fail to reject the null hypothesis, you did not find a significant difference.

- Null Hypothesis statement: **H0: μ1=μ2**
- Alternative Hypothesis statement: **HA: μ1≠μ2**

Hypothesis Testing Steps

- **Step 1: State the Null Hypothesis and the Alternative Hypothesis**
  - The NULL HYPOTHESIS, H0, is the statement that there is NO RELATIONSHIP.
    - The null hypothesis is the statement that "there is no significant difference between the two *means*, u1 and u2."
  - The alternative hypothesis, HA, is the statement that there IS A RELATIONSHIP for whatever relationship being tested.
    - Your alternative hypothesis would be that "there is a significant difference between the two *means*, u1 and u2."
- **Step 2: Decide on the Significance Level**
  - **SIGNIFICANCE LEVEL**: A number that is used as the cutoff for how statistically meaningful a probability, equal to or more extreme than what was observed, is.
    - Is a decision criterion that specifies the degree of certainty with which you want to make your judgement of whether or not to reject the null hypothesis.
    - The significance level is the probability that you will mistakenly reject a true null hypothesis based on the sample statistic.
    - The more careful you want to be about not rejecting a true null hypothesis, the smaller your significance level should be:
      - A **higher significance level** indicates a **higher threshold to reject the null hypothesis**. To state that there is a significant difference, you have to be more certain that random chance or error is not causing the difference.
      - A **lower significance level** indicates a **lower threshold to reject the null hypothesis**. To state that there is a significant difference,

you do not have to be as certain that random chance or error is not causing the difference.

    - A commonly used significance level in many research settings is 0.05.
        - Another way of saying this is that if you find a significant result (i.e., your obtained p-value is less than .05) then you are 95% confident that you are correct in your decision (i.e.,your obtained result is within the 95% confidence interval).

- **Step 3: Compute the Value of the Test Statistic**
    - **TEST STATISTICS: (t-stat);** One value used to test the hypothesis, it is a numerical summary of the dataset.
        - For example, you may be testing the difference between a sample mean, and a population mean. One of the most useful kinds of test statistics (that can be used for hypothesis testing in this situation) is the One-Sample t-Test. ***The One-Sample t-Test can be used to test a null hypothesis concerning a population mean based on statistics from one random sample from the population.***
        - The test statistic for a One-Sample *t*-test is:

          $$t = \frac{\bar{x} - \mu_0}{s/\sqrt{n}}$$

          where:

          $\bar{x}$ = the sample average
          s = the standard deviation of the sample values
          n = the number of values in the sample
          $\frac{s}{\sqrt{n}} = S_{\bar{x}}$ = standard error of the mean

          To compute the value of the test-statistic, we calculate or identify each of the necessary values: $\bar{x}$, s, n, μ_0. After plugging the appropriate values into the formula, the test statistic value (*t*) is computed using arithmetic. The value of the test statistic is crucial as we move on to the fourth and final step of our hypothesis test.

- **Step 4: Find the Critical Value and Compare to Test Statistic Value**
    - The goal of the hypothesis test is to make a decision about the null hypothesis. This determination is made by comparing the CRITICAL VALUE to the TEST STATISTIC VALUE.
        - **CRITICAL VALUE**: The tipping point between a test statistic value that causes us to reject the null hypothesis and one that indicates we fail to reject the null hypothesis.
        - **TEST STATISTIC**: One value used to test the hypothesis, it is a numerical summary of the dataset.
    - The critical value depends upon the significance level and the test statistic that was employed. Once we calculate the test statistic value, the critical value needs to be found. Having both of these values, we can compare them to one another.
        - If the absolute value of the test statistic is greater than the critical value, we reject the null hypothesis: there is statistical significance. So, the

critical value is the tipping point between where we reject the null hypothesis and where we fail to reject the null hypothesis.

- To determine whether the test statistic's absolute value is large enough to reject the null hypothesis, we must find the critical values for the hypothesis test from a distribution table.
    - DISTRIBUTION TABLE: A table of values used in hypothesis testing to determine the critical value.
- There are two possible outcomes:
    - If the test statistic's value is **NOT** greater than the critical value, we FAIL TO REJECT the null hypothesis. For a t-test, this means is that the difference between the sample mean and μ0 is not significant. There is no significant difference. There is not enough evidence to conclude that the null hypothesis is false.
    - On the other hand, if the test statistic's value **IS** greater than the critical value, we REJECT the null hypothesis. For a t-test, this means is that the difference between the sample mean and μ0 is significant. There is a significant difference. Therefore, there is enough evidence to conclude that the null hypothesis is false.

Notes:

- **T-TEST**: One sample t-test - A hypothesis test that used to compare a sample mean to a known value, often a population mean.
- The goal of the experiment is to determine if the null hypothesis can be rejected.
- For a single sample mean, the alternative hypothesis is the argument that either a sample is not equal to, greater than, or less than the hypothesized null sample.

3.07 Chi-Squared Test

Chi-Squared Test

- **CHI-SQUARED TEST:** A hypothesis test that is used to examine the distribution of categorical data.
    - Commonly used in statistics to draw inferences about a population, by testing sample data.
    - Any hypothesis test has both a null hypothesis* and an alternative hypothesis*. Chi-square tests are no different. The null hypothesis, or H0, is the statement that there is no relationship. The alternative hypothesis, or HA, is the opposite statement to the null hypothesis. It states that there is a relationship for whatever relationship is being tested.
    - Null Hypothesis: states that the data is distributed as expected.
    - Alternative Hypothesis: states that the data is not distributed as expected.

3.08 ANOVA

ANOVA

- **ANALYSIS OF VARIANCE (ANOVA):**
    - A technique used to determine if there is a significant difference among **three or more means.**

        - Using ANOVA, we see if there is sufficient evidence from sample data of three or more populations to determine whether the population means are all equal, or whether there is a significant difference with at least one of the means.
        - The null hypothesis* claims that all population means are equal. For example, if three populations are being tested, the null hypothesis would be H0: $\mu1=\mu2=\mu3$.
        - The alternative hypothesis* states that not all of the population means are equal. We accept the alternative hypothesis if at least one of the population means is considered significantly different.
        - An F-value is the test statistic* that is utilized in ANOVA. As we've seen with other test statistics, our test statistic value and the critical value* determine whether we "reject the null hypothesis" or "fail to reject the null hypothesis." As always, if our test statistic value exceeds the critical value, we reject the null hypothesis. We would conclude that at least one of the population means is significantly different from the others.
        - An F statistic is a value you get when you run an ANOVA test or a regression analysis to find out if the means between two populations are significantly different. It’s similar to a T statistic from a T-Test; A T-test will tell you if a single variable is statistically significant and an F test will tell you if a group of variables are jointly significant.

<u>3.09 Differential Statistical Techniques</u>

- Tools that help business organizations make informed decision making:
    - Regression analysis: a way to measure how one variable is related to another.
        - Commonly used for forecasting future sales, analyzing cost behavior, and anticipating customer patterns.
        - Identifies a function that describes, as closely as possible, the relationship between these variables so that we can predict what value one variable will assume, if we know the specific value of the other one.
    - Time series analysis: s set of evenly spaced numerical observations on a quantitative variable collected over regular time periods.
        - Assumes that factors influencing the past and present outcomes will continue to influence the outcomes in the future.
        - Independent variable
        - Trends, cyclicalities, seasonalities, and irregularities are the four data patterns used to analyze time-series data.

<u>3.10 Forecasting, Regression Analysis and Quantitative Techniques</u>

- Forecasting:
    - Judgemental: based on sales, consumer, or management input
    - Time series: based upon data patterns in past data, which includes techniques for random variation, trend, seasonality, etc.

- Associative: based upon predictive or explanatory variables and includes regression.
- **REGRESSION ANALYSIS**: A statistical analysis tool that quantifies the relationship between a dependent variable and one or more independent variables.
  - Method to measure the average amount of change in a dependent variable associated with a unit change in one or more independent variables.
  - Used for analyzing cost behavior and forecasting future sales.

- Quantitative Techniques:
  - **TIME SERIES ANALYSIS:** Regression analysis that uses time as the independent variable.
    - Commonly used for evaluating patterns in data to make decisions about staffing levels, inventory, etc.
  - **CLUSTER ANALYSIS:** The process of arranging terms or values based on different variables into "natural" groups.
    - Commonly used for understanding the makeup of an industry's different areas.
  - **DECISION ANALYSIS:** The process of weighincg all outcomes of a decision to determine the best course of action.
    - Commonly used in making decisions, whether personal or professional.

3.11 Benefits and Shortcomings of Regression Analysis and Quantitative Analysis

	Benefits	Shortcomings
Regression analysis	- Allows sophisticated analysis of cost behavior and sales forecasts - Provides objective benchmarks for evaluation of reliability of estimates	- Requires 15 or more data points for accuracy - Can be influenced by outliers (unusual data points) - Requires informed analysis
Time series analysis	- Aids decision making by finding patterns in data, such as sales trends - Allows performance and productivity evaluation	- Assumes past data patterns will repeat in future, which may not be true - Key variables may not be captured
Cluster analysis	- Sorts individual data points into different groups - Helps determine target markets - Identifies successful and unsuccessful habits and systems	- Long and expensive process - There are hundreds of potential approaches to take, each specific to a certain situation

Decision analysis	- Determines the decision with the greatest value - Produces a value under certainty, uncertainty, and risk	- Quality of decision is limited to the amount of data available - Does not emphasize the risk of the worst case scenario

3.12 Regression Analysis

- Goal: Understand the function explaining the impact of the independent variable on the dependent variable. Can be used to predict future market values.
- Looks like scatter plots.
- Linear Regression: One independent and one dependent variable.
- Multiple Regression Analysis: MORE than one independent variable and has one dependent variable.
    - Goal: Find an equation to explain the relationship between variables with minimized error.
    - Helps identify which is most strongly correlated with the dependent variable.

3.13 Linear Regression

Regression Analysis

- **REGRESSION ANALYSIS:** A statistical analysis tool that quantifies the relationship between a dependent variable and one or more independent variables.
    - Useful tool when you want to analyze the relationship between two or more variables.
    - If we know that an association exists between two variables, we can use regression for description and prediction. Specifically, we can use regression analysis to describe a trend or predict values based on known values using the regression analysis equation.
- DEPENDENT VARIABLE: The variable whose value depends on one or more variables in the equation; typically the cost or activity to be predicted.
    - Variable whose value “depends” on the other variables in the equation.
- INDEPENDENT VARIABLE: The variables presumed to influence another variable (dependent variable); typically it is the level of activity or cost driver.

Linear Regression

- Positive or negative terms: the direction of the dots on the chart - upwards or downwards.
- Strong or weak terms: how tightly bunched together the dots are.
- A linear relationship between two variables can be measured by its strength.
    - A strong linear relationship indicates that the data will bunch around a straight line, while a weak linear relationship does not. Although it will rarely follow a straight line exactly, if a regression does fall precisely along a straight line, we call it a perfect linear relationship.
    - If a linear relationship exists, it can be either positive or negative. A positive linear relationship exists when both variables increase together. A negative linear

relationship exists when the one variable decreases as the other variable increases.

Correlation

- The strength of a linear relationship can be measured with the correlation coefficient. It is a number between -1 and 1.
- A correlation coefficient that is close to 0 indicates a weak linear relationship, while a correlation coefficient closer to -1 or 1 represents a strong linear relationship.
- A correlation coefficient equal to exactly -1 or 1 would be considered perfectly linear. In the case of a perfectly linear relationship, our data points would form a perfectly straight line.
- Negative linear relationships have correlation coefficients less than 0.
- Positive linear relationships have correlation coefficients greater than 0.

The R2 (Squared) Statistic

- "R-squared" or "R2" provides a measure of "goodness of fit."
- Ranges from 0 to 1.
    - Value close to 1 indicates that the estimation error is small and data closely aligns to the regression line.

Standard Error (SE) of Estimate

Notes:

- P Value: probability of an observed result assuming that the null hypothesis is true.
    - P value greater than significant value than accept the null hypothesis
    - P value less than significant value than reject the null hypothesis
-