Exercise Explanations

Answer for Exercise 07.pdf

Answer for Exercise 06.pdf

Exercise 05 - Answers.pdf

Exercise 04 - Answers.pdf

Exercise 03 - Answers.pdf

Exercise 02 Answer.pdf

Exercise 01 Answers.pdf

Exercise 01

1. **Create a Stem-and-Leaf Display**
   - **Concept:** This exercise practices *data visualization* using a stem-and-leaf plot.

- **Explanation:**
    - A stem-and-leaf plot is a way to organize and display numerical data.
    - The "stem" represents the leading digit(s), and the "leaf" represents the trailing digit of each data point.
    - This is an easy way to see the distribution, range and clusters of data in a dataset.
    - **Example:** In the provided solution, "6 | 2 5 8" means that the dataset contains the values 62, 65, and 68.
- **Key Points:**
    - It preserves the original data values (unlike histograms).
    - Useful for small to medium-sized datasets.
    - Helps see the shape of the distribution, outliers, and clusters.

2. **Construct a Box Plot**
    - **Concept:** This exercise practices creating a *box plot*, a way of visualizing the distribution of data using the five-number summary.
    - **Explanation:**
        - A box plot (or box-and-whisker plot) is a standardized way of displaying data based on a five-number summary.
        - **Five-Number Summary:**
            - **Minimum:** The smallest value in the dataset.
            - **First Quartile (Q1 or 25th percentile):** The median of the lower half of the data.
            - **Median (Q2 or 50th percentile):** The middle value of the sorted data.
            - **Third Quartile (Q3 or 75th percentile):** The median of the upper half of the data.
            - **Maximum:** The largest value in the dataset.
        - **Interquartile Range:** The IQR is Q3-Q1.

    - **Box Plot Components:**
        - **Box:** Spans from Q1 to Q3.
        - **Median Line:** A line within the box represents the median.
        - **Whiskers:** Lines that extend out from the box, to find the adjacent values. A maximum of 1.5*IQR.
        - **Outliers:** The points that lie beyond the whiskers are considered outliers.
    - **Calculating Outliers:**
        - **Upper Inner Fence:** Q3 + 1.5 * IQR
        - **Lower Inner Fence:** Q1 - 1.5 * IQR
        - **Upper Adjacent:** Max point in the dataset that is not above the upper inner fence.
        - **Lower Adjacent:** Min point in the dataset that is not below the lower inner fence.
- **Key Points:**
    - Helps visualize the spread and skewness of the data.
    - Easy to compare the distribution of two or more groups by putting them side-by-side.
    - Identifies potential outliers.
    - Understanding percentiles and the 5-number summary is key.

Exercise 02

1. **Calculate the Trimean**
    - **Concept:** This exercise introduces the *trimean*, a robust measure of central tendency.
    - **Explanation:**
        - The trimean is a weighted average of the median and the first and third quartiles (25th and 75th percentiles).

        - **Formula:** Trimean = (Q1 + 2 * Median + Q3) / 4
        - **Why Use It?**
            - It's more robust than the mean to the presence of outliers because it gives less weight to extreme values.
            - It combines information from the median and the quartiles for a more stable center representation.
        - **Calculation:** The exercise shows how to calculate the median, Q1, and Q3, and then plug them into the trimean formula.
    - **Key Points:**
        - Understand the five-number summary for calculating the trimean.
        - Useful as an alternative to the mean for skewed data or with possible outliers.
2. **Geometric Mean**
    - **Concept:** This exercise shows the *geometric mean*, used for finding the average rate of change or growth over time.
    - **Explanation:**
        - The geometric mean is a type of average used when dealing with rates, ratios, or percentages.
        - It is calculated by multiplying all the numbers together and then taking the nth root, where n is the number of values.
        - **Formula:** GM = $\sqrt[n]{(x1 * x2 * ... * xn)}$
        - **Why Use It?**
            - It's appropriate for growth rates because it avoids the distortion caused by arithmetic means when dealing with multiplicative changes.
            - Useful for financial returns, population growth, or other time-series data.
        - **Calculation:** This problem outlines how to turn percentages into growth factors and then apply the equation.

- **Key Points:**
    - Recognize when a geometric mean is appropriate over arithmetic mean.
    - Be able to turn percentages into growth factors.

3. **Trimmed Mean**
    - **Concept:** This exercise focuses on the *trimmed mean*, another measure of central tendency that reduces outlier effects.
    - **Explanation:**
        - A trimmed mean is calculated by removing a specified percentage of data points from both the lower and upper ends of a sorted dataset. Then finding the average of the remaining values.
        - **Why Use It?**
            - It's more robust to extreme values than the mean because it excludes outliers.
            - It aims to provide a more representative "typical" value when your data may contain errors or unusual cases.
        - **Calculation:** The solution shows how to sort the data, trim the specified percentage of values from each end, and calculate the mean of the remaining values.
    - **Key Points:**
        - Understanding how to calculate percentages from the dataset.
        - Useful to reduce outliers effect on average.

Exercise 03

1. **Permutations**
    - **Concept:** This exercise introduces *permutations*, an important topic in counting.
    - **Explanation:**

    - A permutation is a way of arranging items where the order of items **matters**.
    - **Formula:** P(n, r) = n! / (n - r)! , where n is the number of total items, and r is the number of items to select.
    - **Example:** In the given problem, choosing 4 people out of 8 for a photo where the order they stand in a row is important, then we need to use permutation.
    - **Calculation:** The solution outlines how to calculate P(8,4) using the formula.
- **Key Points:**
    - Recognize when to use permutation (order matters) instead of combination.
    - Comfortable with factorials (!).

2. **Combinations**
    - **Concept:** This exercise uses *combinations*, the opposite of permutations in counting.
    - **Explanation:**
        - A combination is a way of selecting items where the order of the items **does not matter**.
        - **Formula:** C(n, r) = n! / (r! * (n - r)!), where n is the number of total items and r is the number of items to select.
        - **Example:** When selecting 4 books out of 7 for a trip, the order does not matter (just the chosen four).
        - **Calculation:** The solution shows how to calculate C(7, 4) using the formula.
    - **Key Points:**
        - Recognize when to use combination (order doesn't matter) versus permutations.
        - Understand the difference between selecting and arranging.

3. **Hypergeometric Probability,** https://stattrek.com/online-calculator/hypergeometric
   - **Concept:** This problem uses the *hypergeometric distribution* to calculate probabilities.
   - **Explanation:**
     - The hypergeometric distribution calculates the probability of successes (specific outcomes) in a set number of trials *without replacement*.
     - **Formula:** P(X=k) = ((kCx)(N-k)C(n-x))/ (NCn), where N is the population size, k is the number of successes, n is number of trials, x is number of successes.
     - **Example:** Here we calculate the probability of drawing *exactly* 3 red balls out of a selection of 5.
     - **Calculation:** This problem breaks down the application of the formula in steps.
   - **Key Points:**
     - Know when the Hypergeometric distribution applies - sampling without replacement.
     - Understand how it differs from the binomial distribution.

Exercise 04

1. **Geometric Mean**
   - This is identical to the same concept already discussed in Exercise 02, #2
2. **Box Plot Comparison**
   - This exercise is identical to what was already discussed in Exercise 01 #2
3. **Probability of Independent Events**
   - **Concept:** This exercise deals with the probability of *independent events*, and how to find the probability of those happening together.
   - **Explanation:**

    - Two events are independent if the outcome of one event does not affect the outcome of the other event.
    - To find the probability of two independent events happening, you multiply their individual probabilities.
    - **Formula:** P(A and B) = P(A) * P(B)
    - **Example:** This exercise uses a coin flip and a deck of cards as an example.
  - **Key Points:**
    - Recognizing whether events are independent or dependent.
    - Understand how to find the probability of multiple independent events occuring.

4. **Back-to-Back Stem-and-Leaf Plot**
   - **Concept:** This problem uses a *back-to-back stem-and-leaf plot* which compares to data sets.
   - **Explanation:**
     - This is a variation of the stem-and-leaf plot, where two sets of values are compared by sharing a common stem and using leaves on either side.
     - Allows to compare the shape, spread and central tendencies of the datasets.
   - **Key points:**
     - Easy to use to visualise and compare two different datasets.
5. **Binomial Probability**
   - **Concept:** This problem is an application of the *binomial distribution*.
   - **Explanation:**
     - The binomial distribution is used for binary outcomes in a set number of independent trials.

        - **Formula:** $P(X=k) = (nCk)(p^k)(1-p)^{n-k}$, where n is the number of trials, k is the number of successes, and p is the probability of success.
    - **Key Points**
        - Understand the conditions of using the binomial distribution, binary outcomes, independent trials, and a fixed number of trials.

6. **Probability of multiple Binomial Events**
    - **Concept:** This builds on the previous binomial probability problem to calculate at least a certain number of successful trials.
    - **Explanation:**
    - You need to use the binomial distribution to calculate the probability of each case individually.
    - For a probability of at least n, you need to add the probability for the case of n, n+1, n+2 up to the max.
        - **Key Points:**
        - Know how to interpret greater than or less than binomial trials to calculate probability.
7. **Pearson Correlation Coefficient**
    - **Concept:** This exercise focuses on calculating the *Pearson correlation coefficient*, a measure of linear association.
    - **Explanation:**
        - The Pearson correlation coefficient measures the strength and direction of a linear relationship between two continuous variables.
        - **Range:** -1 to +1.
            - +1 is a perfect positive linear correlation.
            - 1 is a perfect negative linear correlation.
            - 0 indicates no linear correlation.
    - **Formula:** r = cov(X,Y)/ (SDx * SDy)

- **Calculation:** This exercise demonstrates the steps of calculating the correlation based on the equation.
- **Key Points:**
    - Understand the range of correlation values.
    - Be able to describe the strength and direction of relationships.
    - Important to understand that correlation does not equal causation.

Exercise 05

1. **Standard Deviation**
    - **Concept:** This exercise focuses on calculating the *standard deviation*, a measure of data dispersion.
    - **Explanation:**
        - Standard deviation is a measure of how spread out numbers are from the mean.
        - It measures how much the data deviates on average from the mean.
        - **Calculation**
            1. Calculate the mean.
            2. Find the deviation of each value from the mean.
            3. Square each deviation.
            4. Find the average of the squared deviations (variance).
            5. Take the square root of the variance.
    - **Key Points:**
        - Understanding that the variance is used to calculate the standard deviation.
2. **Z-score & Normal Approximation to the Binomial Distribution**
    - **Concept:** This exercise applies the *normal approximation* to the binomial distribution using z-scores.
    - **Explanation:**

    - When the number of trials (n) in a binomial distribution is large enough, a normal distribution can approximate the distribution.
    - The z-score is used to map from the binomial distribution into a standard normal distribution.
    - To use a normal approximation, the following conditions must be satisfied:
        - $n * p > 5$
        - $n *(1-p) > 5$
    - **Calculation:** The solution outlines how to check the conditions, calculate the mean and standard deviation of the binomial, and then apply the z-score formula. A correction for continuity is also applied.
    - **Formula:** $Z = (X - \mu) / \sigma$
- **Key Points:**
    - Recognize when a normal approximation is appropriate.
    - Be comfortable with standardizing the score using z-scores.
    - Know how to use the normal distribution table.

3. **Normal Approximation to the Binomial Distribution**
    - This is the same concept used in the previous question but using at least as many successes.

Exercise 06

https://www.meracalculator.com/math/t-distribution-critical-value-table.php

1. **One-Sample t-test, in google sheet**
    - **Concept:** This practices a *one-sample t-test* which compares a sample mean against a known population mean.
    - **Explanation:**
        - A one-sample t-test is used when you have a sample, and you want to test if the sample mean is significantly different from a known

population mean (or a hypothesized mean), when the population standard deviation is unknown.

- **Hypotheses:**
  - **Null Hypothesis (H0):** There is no difference between the sample and population means (μ = specified value).
  - **Alternative Hypothesis (H1):** There is a difference between the sample and population means ($\mu \neq$ specified value, or $<$ or $>$ depending on whether it's one-tailed or two-tailed).
- **T-Statistic:** $(\bar{x} - \mu) / (s/\sqrt{n})$ where $\bar{x}$ is the sample mean, μ is the population mean, s is the sample standard deviation and n is the sample size.
- **Degrees of Freedom (df):** n - 1, where n is the sample size.
- **P-Value:** Use the t-statistic to find the p-value.
- **Decision:** Compare the p-value with the significance level. If the p-value is below the chosen significance, we reject the null hypothesis.

- **Key Points:**
  - Understand the t-test statistic and how to calculate it.
  - Know how to interpret degrees of freedom and the t-table/p-values.
  - Know when to do a one-tailed versus a two-tailed test.

2. **Paired Samples t-test,** https://www.statskingdom.com/paired-t-test-calculator.html
   - **Concept:** This uses a *paired-samples t-test*, designed for before-and-after or matched-pair data.
   - **Explanation:**
     - A paired-samples t-test is used when you have two measurements for each unit/subject (e.g., before and after treatment, matched pairs).
     - **Hypotheses:**
       - **Null Hypothesis (H0):** There is no difference in the mean of the paired data (average difference = 0).

        - **Alternative Hypothesis (H1):** There is a difference in the mean of the paired data (average difference ≠ 0 or < or > 0).
    - **Difference Scores:** First, you calculate the differences for each pair. Then calculate the average difference (d).
    - **T-Statistic:** $t = d / (sd/\sqrt{n})$ where d is the average difference, sd is the sample standard deviation of the differences, and n is the number of pairs.
    - **Degrees of Freedom (df):** n-1 where n is the number of pairs.
    - **Decision:** Use the test statistic to find the p-value. if the p-value is less than the significance value (alpha), then reject the null hypothesis.
- **Key Points:**
    - Recognize when paired t-tests are appropriate.
    - Understand that you use the *differences* in paired data.

3. **Independence(unpaired) t-test value,** https://www.graphpad.com/quickcalcs/ttest1/?format=SD

https://yuppal.people.ysu.edu/econ_3790/t-table.pdf - Degree of freedom table(1-100 completed)

- For the t-test calculator, follow up with the correct format such as:
    - Choose SD, mean and n : If we have standard deviation, mean and sample size value in the table to calculate stuffs
- **Concept**
    - The independent samples t-test is a statistical hypothesis test used to compare the means of two independent groups. It determines whether there is a statistically significant difference between the average values of a particular variable in two unrelated groups.
    - **Hypotheses:**
        - **Null Hypothesis (H_0):** The means of the two groups are equal. ($\mu_1 = \mu_2$)
        - **Alternative Hypothesis (H_1):**

        - **Two-tailed:** The means of the two groups are not equal. ($\mu_1 \neq \mu_2$)
        - **One-tailed (left-tailed):** The mean of group 1 is less than the mean of group 2. ($\mu_1 < \mu_2$)
        - **One-tailed (right-tailed):** The mean of group 1 is greater than the mean of group 2. ($\mu_1 > \mu_2$)
    - **Key Points:**
        - **Independence:** The two groups being compared must be independent of each other. This means that the observations in one group should not be related to or influence the observations in the other group.
        - **Normality:** The data in each group should be approximately normally distributed. However, the t-test is relatively robust to violations of normality, especially with larger sample sizes.
        - **Homogeneity of variance (optional):** The variances of the two groups should be equal. This assumption can be tested using Levene's test. If the assumption is violated, a modified version of the t-test (Welch's t-test) can be used.

Calculation Process:

1. **Calculate the mean and standard deviation of each group.**
2. **Calculate the pooled variance (if assuming equal variances):**
    - Pooled variance (s^2) = $[(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2] / (n_1 + n_2 - 2)$
    - where:
        - n_1 and n_2 are the sample sizes of the two groups
        - s_1^2 and s_2^2 are the variances of the two groups
3. **Calculate the standard error of the difference between means:**
    - Standard error (SE) = $\sqrt{[s^2(1/n_1 + 1/n_2)]}$ (if equal variances)
    - Standard error (SE) = $\sqrt{[(s_1^2/n_1) + (s_2^2/n_2)]}$ (if unequal variances)
4. **Calculate the t-value:**

- $t = (M_1 - M_2) / SE$
- where:
    - M_1 and M_2 are the means of the two groups

5. **Determine the degrees of freedom (df):**
    - $df = n_1 + n_2 - 2$ (if equal variances)
    - df is approximated using a more complex formula (Welch-Satterthwaite equation) if unequal variances are assumed.
6. **Find the p-value:**
    - Compare the calculated t-value to the t-distribution with the appropriate degrees of freedom.
    - The p-value represents the probability of observing a t-value as extreme or more extreme than the calculated value, assuming the null hypothesis is true.

Interpretation of Results:

- **If p-value < α (significance level):** Reject the null hypothesis. There is a statistically significant difference between the means of the two groups.
- **If p-value ≥ α:** Fail to reject the null hypothesis. There is not enough evidence to conclude that the means of the two groups are different.

Different Scores:

- **Positive t-value:** The mean of group 1 is significantly greater than the mean of group 2.
- **Negative t-value:** The mean of group 1 is significantly less than the mean of group 2.
- **t-value close to zero:** There is no significant difference between the means of the two groups.

In Summary:

The independent samples t-test is a valuable tool for comparing the means of two independent groups. By following the steps outlined above, researchers can determine whether there is a statistically significant difference between the groups and draw meaningful conclusions from their data.

Exercise 07

1. **One-Way ANOVA**
    - **Concept:** This applies a *one-way ANOVA* to test the means of multiple groups.
    - **Explanation:**
        - This problem outlines all of the necessary steps to calculate a one-way ANOVA.
        - **Hypotheses:**
            - **Null Hypothesis (H0):** There is no difference in the means of the dependent variable across all groups.
            - **Alternative Hypothesis (H1):** At least one of the group means is different from the others.
        - **Sums of Squares:** Calculates the total sum of squares, between group sum of squares and within group sum of squares
        - **Mean squares:** Calculates the mean of squares for between and within group by dividing by their respective degrees of freedom.
        - **F-Statistic** Calculates the F-statistic by dividing the mean square of between groups and mean square of within groups.
        - **Decision:** Comparing the F-statistic against the critical value.
        - **Key Points:**
            - Recognize the assumptions underlying the analysis.
            - Understand the test statistics and how to interpret p-values.
            - Be able to formulate conclusions based on results.

2. **Chi-Square Independence Test,** https://www.socscistatistics.com/tests/chisquare2/default2.aspx
   - **Concept:** This uses the chi-squared test of independence to test if two categorical variables are independent of one another.
   - **Explanation:**
     - You are testing if there is a relationship between two variables.
     - **Hypotheses:**
       - **Null Hypothesis (H0):** The two variables are independent.
       - **Alternative Hypothesis (H1):** The two variables are not independent.
     - **Expected Frequencies:** You need to calculate the expected values assuming independence.
     - **Test Statistic:** The calculated value measures how far off the data is from the expected results.
     - **Degrees of Freedom:**
     - **Decision:** Compare the chi-squared statistic against the critical value. If the test statistic is greater than the critical value, reject the null hypothesis.
   - **Key Points:**
     - Know how to formulate expected frequencies and apply the formula.
     - Understand how to find degrees of freedom and interpret values.
     - Understand how to make conclusions.
3. **Two-Way ANOVA test, sheets or** https://www.statskingdom.com/two-way-anova-calculator.html
   - **Core Purpose and When to Use It**
     - **Purpose:** The two-way ANOVA is designed to assess the effects of *two* categorical independent variables (factors) on a single continuous dependent variable *simultaneously*. It's a powerful tool that allows you to explore not only how each factor influences the dependent variable

separately (main effects) but also if they interact with each other (interaction effect).

- **When to Use:**
  - You have *two* categorical factors (e.g., treatment types, group categories, levels of an intervention).
  - You want to see how they *individually* affect a continuous dependent variable (e.g., test scores, measurements, response rates).
  - Critically, you want to investigate if the effects of one factor depend on the levels of another factor (interaction).
  - Examples:
    - Effect of both dosage of a drug (low, medium, high) and the gender (male, female) of patients on their recovery time.
    - Impact of two different teaching strategies and the year level on student performance.

2. Key Concepts

- **Factors (Independent Variables):** Two categorical variables whose impact is being tested. Each factor has two or more levels or groups.
- **Dependent Variable:** The continuous variable being measured or observed.
- **Main Effect:** The effect of *one* factor on the dependent variable, *averaging across the levels of the other factor*. In other words, the effect of one factor "independently".
- **Interaction Effect:** The effect of *both* factors together. An interaction exists when the effect of one factor is *dependent* on the specific levels of the other factor. The effect of factor A changes depending on what factor B is.
- **Cell:** A combination of levels for each factor. For example, if one factor is gender (male or female) and the other factor is education (bachelors, masters, phd), then a specific cell is the combination of "male and bachelor". Each cell should have an equal sample size.

3. Hypotheses in Two-Way ANOVA

- Unlike the one-way ANOVA, a two-way ANOVA has *three* distinct sets of null and alternative hypotheses:
  1. **Main Effect for Factor A (Row Effect):**
     - **Null Hypothesis (H0a):** There is no main effect of Factor A. The means of the dependent variable across the different levels of Factor A are the same, *regardless of the levels of factor B*.
     - **Alternative Hypothesis (H1a):** There *is* a main effect of Factor A; at least one level of Factor A has a different mean from the other levels when averaged over the other variable.
  2. **Main Effect for Factor B (Column Effect):**
     - **Null Hypothesis (H0b):** There is no main effect of Factor B. The means of the dependent variable across the different levels of Factor B are the same, *regardless of the levels of factor A*.
     - **Alternative Hypothesis (H1b):** There *is* a main effect of Factor B.
  3. **Interaction Effect (Factor A x Factor B):**
     - **Null Hypothesis (H0ab):** There is no interaction between Factor A and Factor B. The effect of one factor does *not* depend on the level of the other factor.
     - **Alternative Hypothesis (H1ab):** There *is* an interaction between Factor A and Factor B. The effect of one factor *does* depend on the level of the other factor.

4. How Two-Way ANOVA Works (Simplified):

1. **Total Sum of Squares (SST):**
   - Total variability of data points from the grand mean.
   - SST = Sum (Xij - X) ^2
   - The value from this is always positive, meaning that the farther points are from the mean the bigger the number.
2. **Sums of Squares:** This partitions the total variability into different sources:

- **Sum of Squares for Factor A (SSA):** Variability attributable to differences between the means of Factor A.
- SSB = n_b * sum (Xmean_a -Xgrand)^2
- **Sum of Squares for Factor B (SSB):** Variability attributable to differences between the means of Factor B.
- SSA = n_a * sum (Xmean_b -Xgrand)^2
- **Sum of Squares for the Interaction (SSAB):** Variability attributable to the interaction effect between A and B.
- SSAB = SST - SSA - SSB - SSE
- **Sum of Squares within groups (SSE):** Variability that cannot be accounted for by factor A or B. It is the variation of values around their group's means.
- **Key Idea:** The sum of squares are a way to partition the total variability. SST = SSA + SSB + SSAB + SSE.

3. **Degrees of Freedom:**
    - These reflect how many independent pieces of information are available to estimate a particular parameter.
    - **df for Factor A:** Number of levels in Factor A minus 1 (k-1).
    - **df for Factor B:** Number of levels in Factor B minus 1 (l-1).
    - **df for Interaction:** (levels of Factor A - 1) * (levels of Factor B - 1) = (k-1) * (l-1).
    - **df for Within (Error):** Total sample size - (number of cells) = N - kl where N is the total number of samples, k is the levels of factor A and l is the levels of factor B.
4. **Mean Squares:** Each sum of squares is divided by its corresponding degrees of freedom to get a mean square (e.g., MSA = SSA / dfa).
    - **Mean Square for Factor A (MSA):** MSA = SSA / df(A)
    - **Mean Square for Factor B (MSB):** MSB = SSB / df(B)
    - **Mean Square for Interaction (MSAB):** MSAB = SSAB / df(AxB)

    - **Mean Square Error (MSE):** MSE = SSE / df(E)

5. **Test Statistics (F-ratios):** To test each of the three hypothesis, three separate F-tests are done.
    - **F for Factor A (FA):** MSA / MSE
    - **F for Factor B (FB):** MSB / MSE
    - **F for Interaction (FAB):** MSAB / MSE
6. **P-values:** Each of the F statistics has a p-value, which tells you how consistent the F value is with the null hypothesis.
    - A low p-value means that the evidence is against the null hypothesis, and vice versa.
7. **Decision:** Compare p-values with α (significance level). Reject null if $p \leq \alpha$.

5. Interpreting Results:

- **Significant Main Effects (Low p-value):** Indicates that *at least one* of the levels of that factor has a statistically significant effect on the dependent variable, *averaged across the other factor*. You then need post hoc tests (Tukey or Bonferroni) to see which level is significantly different from other levels.
- **Significant Interaction (Low p-value):** The effect of one factor *depends* on the level of the other factor. This means you can't interpret the main effects in isolation (they're context-dependent).
    - You may have to analyze simple effects.
        - **Simple Effect:** effect of one factor at one level of another factor.
- **Non-Significant Results (High p-value):** Indicates lack of evidence that the factor has a main effect or any interaction effect.
- **P-value** The p-value has to be smaller than the alpha value.
- **Key Take Away:** the p-values will show whether you can reject or fail to reject the null hypothesis.

- **Follow Up tests:** if you reject a null hypothesis, further analysis needs to be done. Such as post hoc tests or simple effect tests.
- **Effect Size:** to see how impactful the change has been you may need to look at the effect size (i.e. partial eta squared).

6. Example Interpretation from the Exercise PDF (Slide 23):

- **No Significant Interaction:** The p-value is 0.311. There is no interaction between sunlight exposure and watering. Therefore the effects of sunlight does not change based on the watering frequency.
- **No Significant Main Effect for Watering Frequency:** The p-value is 0.975. The different levels of watering (Daily vs Weekly) has no significant effect on plant growth.
- **Significant Main Effect for Sunlight Exposure:** The p-value is <0.001. There is a statistically significant effect of different levels of sunlight exposure on plant growth.

7. The ANOVA Table (Slide 27):

- The table includes all the important values for doing the ANOVA calculations.
- Source of Variation - effect being tested.
- df - degree of freedom
- Sum of Squares - variation explained by the effect
- Mean Square - average of sum of squares
- F value - test statistic
- **Key Depth**: Knowing how to calculate the values from a given data is important.

8. Second Example (Slides 28-38)

- A marketing agency wants to test if training method affects sales, and if different salesman performs different.
- This second example shows an application of two-way ANOVA in a sales and performance setting.

- Walks through the steps, showing the calculation process and interpretations.
- Also how to make decisions based on critical values.

Key Distinctions (Important Reminders):

- **Focus:** Two-Way ANOVA focuses on *both* individual factor effects *and* their combined influence (interaction).
- **Number of Factors:** It involves *two* categorical independent variables (one-way only has one).
- **Continuous Data:** Like T-tests, ANOVA looks at continuous dependent variables.
- **Post-Hoc tests and Simple Effect:** Be aware that after a significant interaction or main effect, you may need additional follow up tests.

Let me know if you'd like a deeper dive into any part of this explanation.