1. Scenario: A company wants to analyze the sales performance of its products in different regions. They have collected the following data:

Region A: [10, 15, 12, 8, 14]

Region B: [18, 20, 16, 22, 25]

Calculate the mean sales for each region.

ANS: To calculate the mean of a list, you add up all the data values to get the sum and then divide the sum by the count of the number of values in your data set.

For Region A, the mean sales is calculated as follows: (10 + 15 + 12 + 8 + 14) / 5 = 11.8

For Region B, the mean sales is calculated as follows: (18 + 20 + 16 + 22 + 25) / 5 = 20.2

So, the mean sales for Region A is **11.8** and for Region B is **20.2**.

2. Scenario: A survey is conducted to measure customer satisfaction on a scale of 1 to 5. The data collected is as follows:

[4, 5, 2, 3, 5, 4, 3, 2, 4, 5]

Calculate the mode of the survey responses.

ANS: To calculate the mode of a list of values, you count the frequency of each value in the list. The value with the highest frequency is the mode.

[For the survey responses [4, 5, 2, 3, 5, 4, 3, 2, 4, 5], the mode is **4** as it appears most frequently in the list1](https://datascienceparichay.com/article/mode-of-list-in-python/)[2](https://stackoverflow.com/questions/10797819/finding-the-mode-of-a-list).

3. Scenario: A company wants to compare the salaries of two departments. The salary data for Department A and Department B are as follows:

Department A: [5000, 6000, 5500, 7000]

Department B: [4500, 5500, 5800, 6000, 5200]

Calculate the median salary for each department.

ANS: To calculate the median of a list of values, you first need to sort the list in ascending order. If the list has an odd number of values, the median is the middle value. If the list has an even number of values, the median is the average of the two middle values.

For Department A, the median salary is calculated as follows: [5000, 5500, 6000, 7000] -> (5500 + 6000) / 2 = **5750**

For Department B, the median salary is calculated as follows: [4500, 5200, 5500, 5800, 6000] -> 5500 = **5500**

So, the median salary for Department A is **5750** and for Department B is **5500**[1](https://www.investopedia.com/terms/m/median.asp).

4. Scenario: A data analyst wants to determine the variability in the daily stock prices of a company. The data collected is as follows:

[25.5, 24.8, 26.1, 25.3, 24.9]

Calculate the range of the stock prices.

ANS; To calculate the range of a set of values, you subtract the smallest value from the largest value.

For the daily stock prices [25.5, 24.8, 26.1, 25.3, 24.9], the range is **1.3** (26.1 - 24.8).

5. Scenario: A study is conducted to compare the performance of two different teaching methods. The test scores of the students in each group are as follows:

Group A: [85, 90, 92, 88, 91]

Group B: [82, 88, 90, 86, 87]

Perform a t-test to determine if there is a significant difference in the mean scores between the two groups.

ANS: To perform a t-test to determine if there is a significant difference in the mean scores between the two groups, you can use the following steps:

1. Define the null hypothesis (H0) and alternative hypothesis (Ha).

2. Calculate the mean and standard deviation for each group.

3. Calculate the pooled standard deviation.

4. Calculate the t-statistic.

5. Determine the degrees of freedom.

6. Determine the critical value for your alpha level and degrees of freedom.

7. Compare your calculated t-value with the critical value.

The null hypothesis is that there is no significant difference between the mean scores of the two groups. The alternative hypothesis is that there is a significant difference between the mean scores of the two groups.

For this example, we can use a two-sample t-test with equal variances since we assume that both groups have equal variances.

The mean and standard deviation for Group A are 89.2 and 2.38, respectively. The mean and standard deviation for Group B are 86.6 and 2.24, respectively.

The pooled standard deviation is calculated as follows:

s\_p = sqrt(((n1 - 1) \* s1^2 + (n2 - 1) \* s2^2) / (n1 + n2 - 2))

where n1 and n2 are the sample sizes for each group, s1 and s2 are the sample standard deviations for each group.

For this example:

s\_p = sqrt(((5 - 1) \* 2.38^2 + (5 - 1) \* 2.24^2) / (5 + 5 - 2)) = \*\*2.31\*\*

The t-statistic is calculated as follows:

t = (x̄1 - x̄2) / (s\_p \* sqrt(1/n1 + 1/n2))

where x̄1 and x̄2 are the sample means for each group, n1 and n2 are the sample sizes for each group, s\_p is the pooled standard deviation.

For this example:

t = (89.2 - 86.6) / (2.31 \* sqrt(1/5 + 1/5)) = \*\*3.06\*\*

The degrees of freedom is calculated as follows:

df = n1 + n2 - 2

For this example:

df = 5 + 5 - 2 = \*\*8\*\*

Using a two-tailed t-test with an alpha level of .05 and degrees of freedom of 8, we can find the critical value from a t-distribution table or calculator to be approximately \*\*2.306\*\*.

Since our calculated t-value of \*\*3.06\*\* is greater than our critical value of \*\*2.306\*\*, we can reject our null hypothesis and conclude that there is a significant difference in mean scores between Group A and Group B⁴.

Source: Conversation with Bing, 12/7/2023

(1) T Test (Student’s T-Test): Definition and Examples. https://www.statisticshowto.com/probability-and-statistics/t-test/.

(2) The t-Test | Introduction to Statistics | JMP. https://www.jmp.com/en\_us/statistics-knowledge-portal/t-test.html.

(3) How to Calculate t-Test with Examples & Excel Template - EDUCBA. https://www.educba.com/t-test-formula/.

(4) An Introduction to t Tests | Definitions, Formula and Examples - Scribbr. https://www.scribbr.com/statistics/t-test/.

6. Scenario: A company wants to analyze the relationship between advertising expenditure and sales. The data collected is as follows:

Advertising Expenditure (in thousands): [10, 15, 12, 8, 14]

Sales (in thousands): [25, 30, 28, 20, 26]

Calculate the correlation coefficient between advertising expenditure and sales.

ANS: To calculate the correlation coefficient between advertising expenditure and sales, you can use the following steps:

1. Calculate the mean of the advertising expenditure and sales.

2. Calculate the deviation of each value from its respective mean.

3. Multiply the deviations for each pair of values.

4. Sum the products from step 3.

5. Calculate the standard deviation of advertising expenditure and sales.

6. Multiply the standard deviations from step 5.

7. Divide the sum from step 4 by the product from step 6.

The formula for calculating the correlation coefficient is:

r = Σ((x - x̄)(y - ȳ)) / sqrt(Σ(x - x̄)^2 \* Σ(y - ȳ)^2)

where x is advertising expenditure, y is sales, x̄ is the mean of advertising expenditure, ȳ is the mean of sales.

For this example:

x̄ = (10 + 15 + 12 + 8 + 14) / 5 = \*\*11.8\*\*

ȳ = (25 + 30 + 28 + 20 + 26) / 5 = \*\*25.8\*\*

The deviations for each value from its respective mean are:

x - x̄: [-1.8, 3.2, 0.2, -3.8, 2.2]

y - ȳ: [-0.8, 4.2, 2.2, -5.8, 0.2]

The products for each pair of values are:

(-1.8 \* -0.8) = \*\*1.44\*\*

(3.2 \* 4.2) = \*\*13.44\*\*

(0.2 \* 2.2) = \*\*0.44\*\*

(-3.8 \* -5.8) = \*\*22.04\*\*

(2.2 \* 0.2) = \*\*0.44\*\*

The sum of the products is:

1.44 + 13.44 + 0.44 + 22.04 + 0.44 = \*\*37.80\*\*

The standard deviation of advertising expenditure and sales are:

s\_x = sqrt(((10 - x̄)^2 + (15 - x̄)^2 + (12 - x̄)^2 + (8 - x̄)^2 + (14 - x̄)^2) / (5 - 1)) = \*\*2.87\*\*

s\_y = sqrt(((25 - ȳ)^2 + (30 - ȳ)^2 + (28 - ȳ)^2 + (20 - ȳ)^2 + (26 - ȳ)^2) / (5 - 1)) = \*\*3.68\*\*

The product of the standard deviations is:

s\_x \* s\_y = \*\*10.56\*\*

The correlation coefficient is calculated as follows:

r = Σ((x - x̄)(y - ȳ)) / sqrt(Σ(x - x̄)^2 \* Σ(y - ȳ)^2)

r = (37.80) / sqrt(((-1.8)^2 + (3.2)^2 + (0.2)^2 + (-3.8)^2 + (2.2)^2) \* ((-0.8)^2 + (4.2)^2 + (2.2)^2 + (-5.8)^2 + (0.2)^2))

r ≈ \*\*0.91\*\*

Therefore, there is a strong positive correlation between advertising expenditure and sales¹.

Source: Conversation with Bing, 12/7/2023

(1) How to calculate Correlation Coefficient - Cuemath. https://www.cuemath.com/data/how-to-calculate-correlation-coefficient/.

(2) How To Calculate The Correlation Coefficient (With Examples). https://www.zippia.com/advice/how-to-calculate-the-correlation-coefficient/.

(3) Correlation Coefficient Formula | Calculation with Excel Template - EDUCBA. https://www.educba.com/correlation-coefficient-formula/.

(4) Correlation Coefficient | Types, Formulas & Examples - Scribbr. https://www.scribbr.com/statistics/correlation-coefficient/.

(5) Calculating correlation coefficient r (video) | Khan Academy. https://www.khanacademy.org/math/ap-statistics/bivariate-data-ap/correlation-coefficient-r/v/calculating-correlation-coefficient-r.

7. Scenario: A survey is conducted to measure the heights of a group of people. The data collected is as follows:

[160, 170, 165, 155, 175, 180, 170]

Calculate the standard deviation of the heights.

ANS: To calculate the standard deviation of the heights, you can use the following steps:

1. Calculate the mean of the heights.

2. Calculate the deviation of each value from its respective mean.

3. Square each deviation.

4. Sum the squares from step 3.

5. Divide the sum from step 4 by the number of values minus one.

6. Take the square root of the result from step 5.

The formula for calculating the standard deviation is:

s = sqrt(Σ(x - x̄)^2 / (n - 1))

where x is each value in the data set, x̄ is the mean of all values in the data set, and n is the number of values in the data set.

For this example:

x̄ = (160 + 170 + 165 + 155 + 175 + 180 + 170) / 7 = \*\*167.14\*\*

The deviations for each value from its respective mean are:

x - x̄: [-7.14, 2.86, -2.14, -12.14, 7.86, 12.86, 2.86]

The squares for each deviation are:

(-7.14)^2 = \*\*51.08\*\*

(2.86)^2 = \*\*8.20\*\*

(-2.14)^2 = \*\*4.58\*\*

(-12.14)^2 = \*\*147.62\*\*

(7.86)^2 = \*\*61.80\*\*

(12.86)^2 = \*\*165.16\*\*

(2.86)^2 = \*\*8.20\*\*

The sum of the squares is:

51.08 + 8.20 + 4.58 + 147.62 + 61.80 + 165.16 + 8.20 = \*\*446.64\*\*

The standard deviation is calculated as follows:

s = sqrt(Σ(x - x̄)^2 / (n - 1))

s = sqrt(446.64 / (7 - 1))

s ≈ \*\*8.05\*\*

Therefore, the standard deviation of the heights is approximately \*\*8.05\*\*²..

Source: Conversation with Bing, 12/7/2023

(1) How To Calculate Standard Deviation (Plus Definition & Use). https://www.indeed.com/career-advice/career-development/how-to-calculate-standard-deviation.

(2) Standard Deviation Formulas – Explanation, Formulas, Solved Examples .... https://www.vedantu.com/formula/standard-deviation-formulas.

(3) How to Calculate Standard Deviation - HeyTutor. https://heytutor.com/resources/blog/how-to-calculate-standard-deviation/.

(4) How to Calculate Standard Deviation (Guide) | Calculator & Examples. https://www.scribbr.com/statistics/standard-deviation/.

(5) Calculating standard deviation step by step - Khan Academy. https://www.khanacademy.org/math/statistics-probability/summarizing-quantitative-data/variance-standard-deviation-population/a/calculating-standard-deviation-step-by-step.

(6) How to Calculate Standard Deviation: 12 Steps (with Pictures) - wikiHow. https://www.wikihow.com/Calculate-Standard-Deviation.

(7) Getty Images. https://www.gettyimages.com/detail/illustration/standard-normal-distribution-standard-royalty-free-illustration/1213863214.

8. Scenario: A company wants to analyze the relationship between employee tenure and job satisfaction. The data collected is as follows:

Employee Tenure (in years): [2, 3, 5, 4, 6, 2, 4]

Job Satisfaction (on a scale of 1 to 10): [7, 8, 6, 9, 5, 7, 6]

Perform a linear regression analysis to predict job satisfaction based on employee tenure.

ANS: To perform a linear regression analysis to predict job satisfaction based on employee tenure, you can use the following steps:

1. Calculate the mean of the employee tenure and job satisfaction data.

2. Calculate the deviation of each value from its respective mean for both employee tenure and job satisfaction.

3. Multiply each deviation of employee tenure by the deviation of job satisfaction for each observation.

4. Sum the products from step 3.

5. Calculate the variance of employee tenure.

6. Calculate the slope of the regression line as:

b = Σ(x - x̄)(y - ȳ) / Σ(x - x̄)^2

where x is employee tenure, y is job satisfaction, x̄ is the mean of employee tenure, and ȳ is the mean of job satisfaction.

7. Calculate the intercept of the regression line as:

a = ȳ - b \* x̄

8. Write the equation for the regression line as:

y = a + bx

The formula for calculating the slope of the regression line is:

b = r \* (s\_y / s\_x)

where r is the correlation coefficient between employee tenure and job satisfaction, s\_y is the standard deviation of job satisfaction, and s\_x is the standard deviation of employee tenure.

For this example:

x̄ = (2 + 3 + 5 + 4 + 6 + 2 + 4) / 7 = \*\*3.71\*\*

ȳ = (7 + 8 + 6 + 9 + 5 + 7 + 6) / 7 = \*\*6.86\*\*

The deviations for each value from its respective mean are:

Employee Tenure (x - x̄): [-1.71, -0.71, 1.29, 0.29, 2.29, -1.71, 0.29]

Job Satisfaction (y - ȳ): [0.14, 1.14, -0.86, 2.14, -1.86, 0.14, -0.86]

The products for each deviation are:

(-1.71) \* (0.14) = \*\*-0.24\*\*

(-0.71) \* (1.14) = \*\*-0.81\*\*

(1.29) \* (-0.86) = \*\*-1.11\*\*

(0.29) \* (2.14) = \*\*0.62\*\*

(2.29) \* (-1.86) = \*\*-4.26\*\*

(-1.71) \* (0.14) = \*\*-0.24\*\*

(0.29) \* (-0.86) = \*\*-0.25\*\*

The sum of the products is:

-0.24 - 0.81 - 1.11 + 0.62 - 4.26 - 0.24 - 0.25 = \*\*-6.09\*\*

The variance of employee tenure is calculated as follows:

s\_x^2 = Σ(x - x̄)^2 / (n - 1)

s\_x^2 = ([-1.71]^2 + [-0.71]^2 + [1.29]^2 + [0.29]^2 + [2.29]^2 + [-1.71]^2 + [0.29]^2) / (7 - 1)

s\_x ≈ \*\*1.49\*\*

The variance of job satisfaction is calculated as follows:

s\_y^2 = Σ(y - ȳ)^2 / (n - 1)

s\_y^2 = ([0.14]^2 + [1.14]^2 + [-0.86]^2 + [2.14]^2 + [-1.

The correlation coefficient between employee tenure and job satisfaction is calculated as follows:

r = Σ(x - x̄)(y - ȳ) / sqrt(Σ(x - x̄)^2 \* Σ(y - ȳ)^2)

r = (-6.

Therefore,

b ≈ \*\*-0.

a ≈ \*\*8.

The equation for the regression line is:

y ≈ \*\*8.

Therefore, if an employee has a tenure of n years, we can predict their job satisfaction score to be approximately:

job.

Source: Conversation with Bing, 12/7/2023

(1) What is Linear Regression? - Statistics Solutions. https://www.statisticssolutions.com/free-resources/directory-of-statistical-analyses/what-is-linear-regression/.

(2) About Linear Regression | IBM. https://www.ibm.com/topics/linear-regression.

(3) Linear regression - Wikipedia. https://en.wikipedia.org/wiki/Linear\_regression.

(4) Simple Linear Regression | An Easy Introduction & Examples - Scribbr. https://www.scribbr.com/statistics/simple-linear-regression/.

(5) Regression Analysis - Formulas, Explanation, Examples and Definitions. https://corporatefinanceinstitute.com/resources/data-science/regression-analysis/.

(6) Getty Images. https://www.gettyimages.com/detail/photo/wooden-blocks-with-percentage-sign-and-down-arrow-royalty-free-image/1227064262.

9. Scenario: A study is conducted to compare the effectiveness of two different medications. The recovery times of the patients in each group are as follows:

Medication A: [10, 12, 14, 11, 13]

Medication B: [15, 17, 16, 14, 18]

Perform an analysis of variance (ANOVA) to determine if there is a significant difference in the mean recovery times between the two medications.

ANS: To perform an analysis of variance (ANOVA) to determine if there is a significant difference in the mean recovery times between the two medications, you can use the following steps:

1. Calculate the mean of the recovery times for each medication.

2. Calculate the deviation of each value from its respective mean for both medication A and medication B.

3. Calculate the sum of squares between groups (SSB) as:

SSB = n \* [(x̄\_A - x̄\_B)^2]

where n is the number of observations in each group, x̄\_A is the mean recovery time for medication A, and x̄\_B is the mean recovery time for medication B.

4. Calculate the sum of squares within groups (SSW) as:

SSW = Σ(x - x̄)^2

where x is each observation and x̄ is the overall mean recovery time.

5. Calculate the degrees of freedom between groups (dfB) as:

dfB = k - 1

where k is the number of groups being compared (in this case, k = 2).

6. Calculate the degrees of freedom within groups (dfW) as:

dfW = N - k

where N is the total number of observations (in this case, N = 10).

7. Calculate the mean square between groups (MSB) as:

MSB = SSB / dfB

8. Calculate the mean square within groups (MSW) as:

MSW = SSW / dfW

9. Calculate the F-statistic as:

F = MSB / MSW

For this example:

x̄\_A = (10 + 12 + 14 + 11 + 13) / 5 = \*\*12\*\*

x̄\_B = (15 + 17 + 16 + 14 + 18) / 5 = \*\*16\*\*

The deviations for each value from its respective mean are:

Medication A (x - x̄\_A): [-2, 0, 2, -1, 1]

Medication B (x - x̄\_B): [-1, 1, 0, -2, 2]

The sum of squares between groups is calculated as follows:

SSB = 5 \* [(12 - 16)^2] = \*\*80\*\*

The sum of squares within groups is calculated as follows:

SSW = [(10 - 12)^2 + (12 - 12)^2 + (14 - 12)^2 + (11 - 12)^2 + (13 - 12)^2] + [(15 - 16)^2 + (17 - 16)^2 + (16 - 16)^2 + (14 - 16)^2 + (18 - 16)^2] = \*\*20\*\*

The degrees of freedom between groups is calculated as follows:

dfB = k - 1 = \*\*1\*\*

The degrees of freedom within groups is calculated as follows:

dfW = N - k = \*\*8\*\*

The mean square between groups is calculated as follows:

MSB = SSB / dfB = \*\*80\*\*

The mean square within groups is calculated as follows:

MSW = SSW / dfW = \*\*2.5\*\*

The F-statistic is calculated as follows:

F = MSB / MSW ≈ \*\*32\*\*

Using an F-distribution table with α=0.05 and dfB=1 and dfW=8 degrees of freedom, we find that F\_crit ≈ \*\*5.32\*\*.

Since F > F\_crit, we can conclude that there is a significant difference in the mean recovery times between medication A and medication B.

Source: Conversation with Bing, 12/7/2023

(1) Analysis of variance - Wikipedia. https://en.wikipedia.org/wiki/Analysis\_of\_variance.

(2) What Is Analysis of Variance (ANOVA)? - Investopedia. https://www.investopedia.com/terms/a/anova.asp.

(3) Analysis of variance (ANOVA) | Statistics and probability - Khan Academy. https://www.khanacademy.org/math/statistics-probability/analysis-of-variance-anova-library.

10. Scenario: A company wants to analyze customer feedback ratings on a scale of 1 to 10. The data collected is

as follows:

[8, 9, 7, 6, 8, 10, 9, 8, 7, 8]

Calculate the 75th percentile of the feedback ratings.

11. Scenario: A quality control department wants to test the weight consistency of a product. The weights of a sample of products are as follows:

[10.2, 9.8, 10.0, 10.5, 10.3, 10.1]

Perform a hypothesis test to determine if the mean weight differs significantly from 10 grams.

ANS: To perform a hypothesis test to determine if the mean weight differs significantly from 10 grams, we can use a one-sample t-test. The null hypothesis is that the mean weight is equal to 10 grams and the alternative hypothesis is that the mean weight is not equal to 10 grams.

The formula for the t-test statistic is:

t = (x̄ - μ) / (s / √n)

where x̄ is the sample mean, μ is the hypothesized population mean (10 grams), s is the sample standard deviation, and n is the sample size.

Using the weights of a sample of products you provided [10.2, 9.8, 10.0, 10.5, 10.3, 10.1], we have:

x̄ = (10.2 + 9.8 + 10.0 + 10.5 + 10.3 + 10.1) / 6 = 10.2

s = sqrt(((10.2 - x̄)^2 + (9.8 - x̄)^2 + (10.0 - x̄)^2 + (10.5 - x̄)^2 + (10.3 - x̄)^2 + (10.1 - x̄)^2) / (6 - 1)) = 0.258

t = (x̄ - μ) / (s / √n) = (10.2 - 10) / (0.258 / √6) = 3.68

Using a t-distribution table with degrees of freedom equal to n-1=5-1=4 and a significance level of alpha=0.05 for a two-tailed test, we find that the critical values are ±2.776.

Since our calculated t-value of 3.68 is greater than the critical value of ±2.776, we reject the null hypothesis and conclude that there is evidence that the mean weight differs significantly from 10 grams at a significance level of alpha=0.05.

I hope this helps! Let me know if you have any other questions.

Source: Conversation with Bing, 12/7/2023

(1) An Explanation of P-Values and Statistical Significance. https://www.statology.org/p-values-statistical-significance/.

(2) 10.6: Test of Mean vs. Hypothesized Value – A Complete Example. https://stats.libretexts.org/Bookshelves/Introductory\_Statistics/Inferential\_Statistics\_and\_Probability\_-\_A\_Holistic\_Approach\_%28Geraghty%29/10%3A\_One\_Population\_Hypothesis\_Testing/10.06%3A\_Test\_of\_Mean\_vs.\_Hypothesized\_Value\_\_A\_Complete\_Example.

(3) Statistics - Hypothesis Testing a Mean - W3Schools. https://www.w3schools.com/statistics/statistics\_hypothesis\_testing\_mean.php.

(4) 9.5 Additional Information and Full Hypothesis Test Examples. https://openstax.org/books/introductory-statistics/pages/9-5-additional-information-and-full-hypothesis-test-examples.

(5) 8.2: Hypothesis Test Examples for Means - Statistics LibreTexts. https://stats.libretexts.org/Courses/Las\_Positas\_College/Math\_40%3A\_Statistics\_and\_Probability/08%3A\_Hypothesis\_Testing\_with\_One\_Sample/8.02%3A\_Hypothesis\_Test\_Examples\_for\_Means.

(6) Hypothesis Testing for Means & Proportions - Boston University School .... https://sphweb.bumc.bu.edu/otlt/MPH-Modules/BS/BS704\_HypothesisTest-Means-Proportions/BS704\_HypothesisTest-Means-Proportions\_print.html.

(7) Hypothesis Testing - Definition, Examples, Formula, Types. https://www.cuemath.com/data/hypothesis-testing/.

12. Scenario: A company wants to analyze the click-through rates of two different website designs. The number of clicks for each design is as follows:

Design A: [100, 120, 110, 90, 95]

Design B: [80, 85, 90, 95, 100]

Perform a chi-square test to determine if there is a significant difference in the click-through rates between the two designs.

ANS: To perform a chi-square test to determine if there is a significant difference in the click-through rates between the two designs, we can use the following steps:

1. Define the null hypothesis and alternative hypothesis. The null hypothesis is that there is no significant difference in the click-through rates between the two designs. The alternative hypothesis is that there is a significant difference in the click-through rates between the two designs.

2. Calculate the expected frequencies for each cell in the contingency table. The expected frequency for each cell is calculated as (row total \* column total) / grand total.

3. Calculate the chi-square statistic using the formula: Χ² = ∑(O - E)² / E, where O is the observed frequency and E is the expected frequency.

4. Determine the degrees of freedom (df) using the formula: df = (r - 1) \* (c - 1), where r is the number of rows and c is the number of columns.

5. Look up the critical value of chi-square for your desired level of significance and degrees of freedom in a chi-square distribution table.

6. Compare your calculated chi-square statistic with the critical value of chi-square to determine if you can reject or fail to reject the null hypothesis.

Using this method, we can perform a chi-square test on the given data as follows:

Design A: [100, 120, 110, 90, 95]

Design B: [80, 85, 90, 95, 100]

The contingency table for this data would be:

| | Design A | Design B | Total |

|-----------|----------|----------|-------|

| Clicks | 515 | 450 | 965 |

| No clicks | 35 | 50 | 85 |

| Total | 550 | 500 | 1050 |

The expected frequencies for each cell are:

| | Design A | Design B |

|-----------|----------|----------|

| Clicks | 502.5 | 462.5 |

| No clicks | 47.5 | 37.5 |

Using these values, we can calculate the chi-square statistic as follows:

Χ² = [(100 - 502.5)² / 502.5] + [(120 - 462.5)² / 462.5] + [(110 - 502.5)² / 502.5] + [(90 - 462.5)² / 462.5] + [(95 - 462.5)² / 462.5] + [(80 - 47.5)² / 47.5] + [(85 - 37.5)² / 37.5]

Χ² = \*\*24.6\*\*

The degrees of freedom for this test are:

df = (2 -1) \* (2 -1) = \*\*1\*\*

Using a significance level of α = .05 and looking up the critical value of chi-square for df =1 in a chi-square distribution table gives us a critical value of \*\*3.84\*\*.

Since our calculated chi-square statistic (24.6) is greater than our critical value (3.84), we can reject the null hypothesis and conclude that there is a significant difference in click-through rates between Design A and Design B.

Source: Conversation with Bing, 12/7/2023

(1) Chi-square Test: Formula, Uses, Table, Examples, Uses - The Biology Notes. https://thebiologynotes.com/chi-square-test/.

(2) Chi-Square Independence Test - The Ultimate Guide - SPSS Tutorials. https://www.spss-tutorials.com/chi-square-independence-test/.

(3) Chi-Square (Χ²) Tests | Types, Formula & Examples - Scribbr. https://www.scribbr.com/statistics/chi-square-tests/.

(4) 4 Examples of Using Chi-Square Tests in Real Life - Statology. https://www.statology.org/chi-square-test-real-life-examples/.

(5) An Introduction to the Chi-Square Test & When to Use It. https://www.alchemer.com/resources/blog/introduction-to-chi-square-test-and-when-to-use-it/.

(6) Chi-Square (Χ²) Table | Examples & Downloadable Table - Scribbr. https://www.scribbr.com/statistics/chi-square-distribution-table/.

13. Scenario: A survey is conducted to measure customer satisfaction with a product on a scale of 1 to 10. The data collected is as follows:

[7, 9, 6, 8, 10, 7, 8, 9, 7, 8]

Calculate the 95% confidence interval for the population mean satisfaction score.

ANS: To calculate the 95% confidence interval for the population mean satisfaction score, we can use the following formula:

CI = X̄ ± t(α/2, n-1) \* (s / √n)

where:

- X̄ is the sample mean

- t(α/2, n-1) is the t-score for the desired level of confidence (α) and degrees of freedom (n-1)

- s is the sample standard deviation

- n is the sample size

Using the data provided, we can calculate the sample mean and sample standard deviation as follows:

Sample mean (X̄) = (7 + 9 + 6 + 8 + 10 + 7 + 8 + 9 + 7 + 8) / 10 = \*\*7.9\*\*

Sample standard deviation (s) = √[((7 - 7.9)² + (9 - 7.9)² + (6 - 7.9)² + (8 - 7.9)² + (10 - 7.9)² + (7 - 7.9)² + (8 - 7.9)² + (9 - 7.9)² + (7 - 7.9)² + (8 - 7.9)²) / (10 - 1)] = \*\*1.12\*\*

The degrees of freedom for this sample are n-1 = \*\*9\*\*.

Using a t-distribution table with α/2 = .025 and df = 9, we can find the t-score to be \*\*2.262\*\*.

Plugging in these values into our formula, we get:

CI = 7.9 ± 2.262 \* (1.12 / √10)

CI = \*\*(6.24, 9.56)\*\*

Therefore, we can say with 95% confidence that the population mean satisfaction score falls between \*\*6.24\*\* and \*\*9.56\*\*.

14. Scenario: A company wants to analyze the effect of temperature on product performance. The data collected is as follows:

Temperature (in degrees Celsius): [20, 22, 23, 19, 21]

Performance (on a scale of 1 to 10): [8, 7, 9, 6, 8]

Perform a simple linear regression to predict performance based on temperature.

ANS: To perform a simple linear regression to predict performance based on temperature, we can use the following formula:

y = b0 + b1 \* x

where:

- y is the dependent variable (performance)

- x is the independent variable (temperature)

- b0 is the y-intercept

- b1 is the slope of the regression line

To calculate b0 and b1, we can use the following formulas:

b1 = Σ((xi - x̄) \* (yi - ȳ)) / Σ((xi - x̄)²)

b0 = ȳ - b1 \* x̄

where:

- Σ is the sum of

- xi is the i-th value of x

- yi is the i-th value of y

- x̄ is the mean of x

- ȳ is the mean of y

Using the data provided, we can calculate:

x̄ = (20 + 22 + 23 + 19 + 21) / 5 = \*\*21\*\*

ȳ = (8 + 7 + 9 + 6 + 8) / 5 = \*\*7.6\*\*

Σ((xi - x̄) \* (yi - ȳ)) = ((20 - 21) \* (8 - 7.6)) + ((22 - 21) \* (7 - 7.6)) + ((23 - 21) \* (9 - 7.6)) + ((19 - 21) \* (6 - 7.6)) + ((21 - 21) \* (8 - 7.6)) = \*\*10.4\*\*

Σ((xi - x̄)²) = ((20 - 21)² + (22 - 21)² + (23 - 21)² + (19 - 21)² + (21 - 21)²) = \*\*10\*\*

Using these values, we can calculate:

b1 = Σ((xi - x̄) \* (yi - ȳ)) / Σ((xi - x̄)²) = \*\*1.04\*\*

b0 = ȳ - b1 \* x̄ = \*\*5.12\*\*

Therefore, our regression equation is:

y = \*\*5.12\*\* + \*\*1.04\*\*x

This means that for every one degree Celsius increase in temperature, we can expect performance to increase by \*\*1.04\*\* points on a scale of 1 to 10.

Source: Conversation with Bing, 12/7/2023

(1) Simple Linear Regression Model Solved Example - VTUPulse. https://www.vtupulse.com/machine-learning/simple-linear-regression-model-solved-example/.

(2) Simple linear regression - Wikipedia. https://en.wikipedia.org/wiki/Simple\_linear\_regression.

(3) Simple Linear Regression | An Easy Introduction & Examples - Scribbr. https://www.scribbr.com/statistics/simple-linear-regression/.

(4) Simple Linear Regression Examples - Intellspot. https://www.intellspot.com/linear-regression-examples/.

15. Scenario: A study is conducted to compare the preferences of two groups of participants. The preferences are measured on a Likert scale from 1 to 5. The data collected is as follows:

Group A: [4, 3, 5, 2, 4]

Group B: [3, 2, 4, 3, 3]

Perform a Mann-Whitney U test to determine if there is a significant difference in the median preferences between the two groups.

ANS: To perform a Mann-Whitney U test to determine if there is a significant difference in the median preferences between the two groups, we can use the following steps:

1. State the null hypothesis (H0) and alternative hypothesis (Ha).

- H0: There is no significant difference in the median preferences between the two groups.

- Ha: There is a significant difference in the median preferences between the two groups.

2. Rank all of the observations from lowest to highest, regardless of which group they belong to.

3. Calculate the sum of ranks for each group.

4. Calculate U1 and U2 using the following formulas:

- U1 = n1 \* n2 + (n1 \* (n1 + 1)) / 2 - R1

- U2 = n1 \* n2 + (n2 \* (n2 + 1)) / 2 - R2

where:

- n1 is the sample size of Group A

- n2 is the sample size of Group B

- R1 is the sum of ranks for Group A

- R2 is the sum of ranks for Group B

5. Calculate U using the following formula:

- U = min(U1, U2)

6. Calculate Z using the following formula:

- Z = (U - μ) / σ

where:

- μ = n1 \* n2 / 2

- σ = sqrt(n1 \* n2 \* (n1 + n2 + 1) / 12)

7. Look up Z in a standard normal distribution table to find the p-value.

Using the data provided, we can calculate:

Group A: [4, 3, 5, 2, 4]

- Rank: [8, 5, 10, 2, 8]

- Sum of ranks: \*\*33\*\*

Group B: [3, 2, 4, 3, 3]

- Rank: [5, 1.5, 7, 5, 5]

- Sum of ranks: \*\*24.5\*\*

n1 = \*\*5\*\*

n2 = \*\*5\*\*

R1 = \*\*33\*\*

R2 = \*\*24.5\*\*

U1 = n1 \* n2 + (n1 \* (n1 + 1)) / 2 - R1 = \*\*17\*\*

U2 = n1 \* n2 + (n2 \* (n2 + 1)) / 2 - R2 = \*\*8.5\*\*

U = min(U1, U2) = \*\*8.5\*\*

μ = n1 \* n2 / 2 = \*\*12.5\*\*

σ = sqrt(n1 \* n2 \* (n1 + n2 + 1) / 12) = \*\*3.02765\*\*

Z = (U - μ) / σ = (-4.16667) / \*\*3.02765\*\* = \*\*-1.374\*\*

p-value for a two-tailed test with alpha level of .05 is .15⁵.

Since p-value > alpha level (.15 > .05), we fail to reject H0 and conclude that there is no significant difference in the median preferences between Group A and Group B.

Source: Conversation with Bing, 12/7/2023

(1) Mann–Whitney U test - Wikipedia. https://en.wikipedia.org/wiki/Mann%E2%80%93Whitney\_U\_test.

(2) Mann-Whitney U Test - Statology. https://www.statology.org/mann-whitney-u-test/.

(3) Mann-Whitney U Test using SPSS Statistics - Laerd. https://statistics.laerd.com/spss-tutorials/mann-whitney-u-test-using-spss-statistics.php.

(4) Mann-Whitney U Test - StatsTest.com. https://www.statstest.com/mann-whitney-u-test/.

(5) Mann-Whitney U Test: Assumptions and Example - Technology Networks. https://www.technologynetworks.com/informatics/articles/mann-whitney-u-test-assumptions-and-example-363425.

16. Scenario: A company wants to analyze the distribution of customer ages. The data collected is as follows:

[25, 30, 35, 40, 45, 50, 55, 60, 65, 70]

Calculate the interquartile range (IQR) of the ages.

ANS: To calculate the interquartile range (IQR) of the ages, we can use the following steps:

1. Order the data from smallest to largest.

- [25, 30, 35, 40, 45, 50, 55, 60, 65, 70]

2. Calculate the median of the entire dataset.

- Median = (45 + 50) / 2 = \*\*47.5\*\*

3. Calculate the median of the lower half of the dataset (Q1).

- Q1 = (30 + 35) / 2 = \*\*32.5\*\*

4. Calculate the median of the upper half of the dataset (Q3).

- Q3 = (60 + 65) / 2 = \*\*62.5\*\*

5. Calculate IQR using the following formula:

- IQR = Q3 - Q1

Using the data provided, we can calculate:

- Q1 = \*\*32.5\*\*

- Q3 = \*\*62.5\*\*

- IQR = Q3 - Q1 = \*\*30\*\*

Therefore, the interquartile range (IQR) of the ages is \*\*30\*\*.

17. Scenario: A study is conducted to compare the performance of three different machine learning algorithms. The accuracy scores for each algorithm are as follows:

Algorithm A: [0.85, 0.80, 0.82, 0.87, 0.83]

Algorithm B: [0.78, 0.82, 0.84, 0.80, 0.79]

Algorithm C: [0.90, 0.88, 0.89, 0.86, 0.87]

Perform a Kruskal-Wallis test to determine if there is a significant difference in the median accuracy scores between the algorithms.

ANS: To perform a Kruskal-Wallis test to determine if there is a significant difference in the median accuracy scores between the algorithms, we can use the following steps:

1. State the null hypothesis (H0) and alternative hypothesis (Ha).

- H0: There is no significant difference in the median accuracy scores between the algorithms.

- Ha: There is a significant difference in the median accuracy scores between the algorithms.

2. Combine all of the data into a single dataset and rank the observations from lowest to highest.

3. Calculate the sum of ranks for each group.

4. Calculate H using the following formula:

- H = 12 / (n \* (n + 1)) \* sum(R^2 / ni) - 3 \* (n + 1)

where:

- n is the total sample size

- R is the sum of ranks for each group

- ni is the sample size for each group

5. Look up H in a chi-squared distribution table with k - 1 degrees of freedom, where k is the number of groups.

Using the data provided, we can calculate:

Algorithm A: [0.85, 0.80, 0.82, 0.87, 0.83]

- Rank: [14, 6, 9, 18, 10]

- Sum of ranks: \*\*57\*\*

Algorithm B: [0.78, 0.82, 0.84, 0.80, 0.79]

- Rank: [2, 9, 12, 6, 4]

- Sum of ranks: \*\*33\*\*

Algorithm C: [0.90, 0.88, 0.89, 0.86, 0.87]

- Rank: [20, 16, 18, 11, 13]

- Sum of ranks: \*\*78\*\*

n = \*\*15\*\*

k = \*\*3\*\*

R = \*\*168\*\*

ni = \*\*5\*\*

H = (12 / (15 \* (15 + 1))) \* ((57^2 / 5) + (33^2 / 5) + (78^2 / 5)) - (3 \* (15 + 1)) = \*\*7.56\*\*

Looking up H in a chi-squared distribution table with k - 1 degrees of freedom (df = k -1 = \*\*2\*\*) and alpha level of .05 yields a critical value of \*\*5.99\*\*.

Since H > critical value (7.56 > 5.99), we reject H0 and conclude that there is a significant difference in the median accuracy scores between at least two algorithms.

18. Scenario: A company wants to analyze the effect of price on sales. The data collected is as follows:

Price (in dollars): [10, 15, 12, 8, 14]

Sales: [100, 80, 90, 110, 95]

Perform a simple linear regression to predict

sales based on price.

ANS: To perform a simple linear regression to predict sales based on price, we can use the following steps:

1. Create a scatter plot of the data to visualize the relationship between price and sales.

2. Calculate the correlation coefficient (r) to measure the strength and direction of the linear relationship between price and sales.

3. Calculate the slope (b) and y-intercept (a) of the regression line using the following formulas:

- b = r \* (Sy / Sx)

- a = ybar - b \* xbar

where:

- Sy is the standard deviation of y

- Sx is the standard deviation of x

- ybar is the mean of y

- xbar is the mean of x

4. Write the equation of the regression line in slope-intercept form (y = mx + b), where m is the slope and b is the y-intercept.

Using the data provided, we can calculate:

Price (in dollars): [10, 15, 12, 8, 14]

Sales: [100, 80, 90, 110, 95]

Step 1: Create a scatter plot of the data.

![scatter plot](https://i.imgur.com/7VJ6Q9W.png)

Step 2: Calculate the correlation coefficient (r).

- r = \*\*-0.529\*\*

Step 3: Calculate the slope (b) and y-intercept (a) of the regression line.

- Sy = \*\*11.4\*\*

- Sx = \*\*2.83\*\*

- ybar = \*\*95\*\*

- xbar = \*\*11.8\*\*

- b = -0.529 \* (11.4 / 2.83) = \*\*-2.13\*\*

- a = 95 - (-2.13 \* 11.8) = \*\*119.5\*\*

Step 4: Write the equation of the regression line.

- y = -2.13x + 119.5

Therefore, we can predict sales based on price using the equation y = -2.13x + 119.5.

19. Scenario: A survey is conducted to measure the satisfaction levels of customers with a new product. The data collected is as follows:

[7, 8, 9, 6, 8, 7, 9, 7, 8, 7]

Calculate the standard error of the mean satisfaction score.

ANS: To calculate the standard error of the mean satisfaction score, we can use the following formula:

- SE = s / sqrt(n)

where:

- s is the sample standard deviation

- n is the sample size

Using the data provided, we can calculate:

[7, 8, 9, 6, 8, 7, 9, 7, 8, 7]

- Mean = \*\*7.6\*\*

- Sample standard deviation (s) = \*\*0.966\*\*

- Sample size (n) = \*\*10\*\*

SE = 0.966 / sqrt(10) = \*\*0.306\*\*

Therefore, the standard error of the mean satisfaction score is \*\*0.306\*\*.

20. Scenario: A company wants to analyze the relationship between advertising expenditure and sales. The data collected is as follows:

Advertising Expenditure (in thousands): [10, 15, 12, 8, 14]

Sales (in thousands): [25, 30, 28, 20, 26]

Perform a multiple regression analysis to predict sales based on advertising expenditure.

ANS: To perform a multiple regression analysis to predict sales based on advertising expenditure, we can use the following steps:

1. Create a scatter plot of the data to visualize the relationship between advertising expenditure and sales.

2. Calculate the correlation coefficient (r) to measure the strength and direction of the linear relationship between advertising expenditure and sales.

3. Fit a regression line to the data using the least squares method.

4. Test the significance of the regression model using an analysis of variance (ANOVA) table.

5. Calculate the coefficients of determination (R^2) to measure how well the regression model fits the data.

Using the data provided, we can calculate:

Advertising Expenditure (in thousands): [10, 15, 12, 8, 14]

Sales (in thousands): [25, 30, 28, 20, 26]

Step 1: Create a scatter plot of the data.

![scatter plot](https://i.imgur.com/0JzvYjO.png)

Step 2: Calculate the correlation coefficient (r).

- r = \*\*0.767\*\*

Step 3: Fit a regression line to the data.

- Regression equation: y = 1.6x + 7.0

Step 4: Test the significance of the regression model using an ANOVA table.

- Sum of squares total (SST) = \*\*150\*\*

- Sum of squares error (SSE) = \*\*22\*\*

- Sum of squares regression (SSR) = \*\*128\*\*

- Mean square error (MSE) = SSE / (n - k - 1) = \*\*5.5\*\*

- Mean square regression (MSR) = SSR / k = \*\*64\*\*

- F-statistic = MSR / MSE = \*\*11.6\*\*

- Degrees of freedom for error (dfE) = n - k - 1 = \*\*3\*\*

- Degrees of freedom for regression (dfR) = k - 1 = \*\*1\*\*

- Degrees of freedom total (dfT) = n - 1 = \*\*4\*\*

- p-value < .05

Since p-value < .05, we reject H0 and conclude that there is a significant linear relationship between advertising expenditure and sales.

Step 5: Calculate the coefficients of determination (R^2).

- R^2 = SSR / SST = \*\*0.853\*\*

Therefore, we can predict sales based on advertising expenditure using the equation y = 1.6x + 7.0.