**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.**

Mean sales of Region A => sum of elements / number of elements =>

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

Mean sales of Region B => sum of elements / number of elements =>

(18+20+16+22+25)/5 = 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.**

Mode of the survey response => most frequent element => here it is a multi modal [4,5]

**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.**

Median Salary for Department A

Step 1: Sort the elements in Ascending order

[5000,5500,6000,7000]

Step 2: Check the number of elements are even or not

Here it is even

Step3: If it is even, then add the middle elements and divide by 2

If it is odd, take the centre element

Here it is even… (5500+6000)/2 = 11500/2 = 5750

Median Salary for Department B

Step 1: Sort the elements in Ascending order

[4500,5200,5500,5800,6000]

Step 2: Check the number of elements is even or not

Here it is Odd

Step3: If it is even, then add the middle elements and divide by 2

If it is odd, take the centre element

Here it is odd… = 5500

**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.**

In statistics, the range is the spread of your data from the lowest to the highest value in the distribution. It is a commonly used measure of variability.

Here the maximum value in the list is 26.1 and the minimum value in the list is 24.8.

So, 26.1 – 24.8 = 1.3

**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.**

Step1: Calculate the mean of each group

Group A Mean = (85+90+92+88+91)/5 = 88.4

Group B Mean = (82+88+90+86+87)/5 = 86.6

Step2: Calculate Variance:

Group A = ((85 - 88.4)^2 + (90 - 88.4)^2 + (92 - 88.4)^2 + (88 - 88.4)^2 + (91 - 88.4)^2) / 4

= 5.2

Group B = ((82 - 86.6)^2 + (88 - 86.6)^2 + (90 - 86.6)^2 + (86 - 86.6)^2 + (87 - 86.6)^2) / 4

= 5.2

Step3: Calculate the pooled variance

Pooled Variance (sP^2) = [(nA - 1) \* sA^2 + (nB - 1) \* sB^2] / (nA + nB - 2)

= [(4 \* 5.2) + (4 \* 5.2)] / (5 + 5 - 2) = 5.2

Step4: Calculate the t value

t-value = (μA - μB) / sqrt(sP^2 \* (1/nA + 1/nB))

= (88.4 - 86.6) / sqrt(5.2 \* (1/5 + 1/5)) = 1.12

Step5: Determine the degrees of freedom

Degrees of Freedom (df) = nA + nB - 2 = 5 + 5 - 2 = 8

Step6: Determine the critical t-value based on the significance level and degrees of freedom

at a significance level of 0.05 with 8 degrees of freedom, the critical t-value is approximately 2.306.

Step7: Compare the calculated t-value with the critical t-value:

If the calculated t-value is greater than the critical t-value, there is a significant difference between the means of the two groups.

If the calculated t-value is less than or equal to the critical t-value, there is no significant difference between the means of the two groups.

In this case, the calculated t-value (1.12) is less than the critical t-value (2.306), indicating that there is **no significant difference in the mean scores between Group A and Group B**.

**6. Scenario: A company wants to analyse 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.**

Let’s consider Advertising Expenditure as X and Sales as Y.

Step1: Calculate mean of X and Y

Mean of X = (10+15+12+8+14)/5 = 11.8

Mean of Y = (25+30+28+20+26)/5 = 25.8

Step2: calculate the standard Deviations of all X and Y from Mean

Deviations for X: [10 - 11.8, 15 - 11.8, 12 - 11.8, 8 - 11.8, 14 - 11.8] = [-1.8, 3.2, 0.2, -3.8, 2.2]

Deviations for Y: [25 - 25.8, 30 - 25.8, 28 - 25.8, 20 - 25.8, 26 - 25.8] = [-0.8, 4.2, 2.2, -5.8, 0.2]

Step3: Calculate the product of the deviations for X and Y

Product of deviations: [-1.8 \* -0.8, 3.2 \* 4.2, 0.2 \* 2.2, -3.8 \* -5.8, 2.2 \* 0.2] = [1.44, 13.44, 0.44, 22.04, 0.44]

Step4: Calculate the sum of product of deviations

Sum of products of deviations = 1.44 + 13.44 + 0.44 + 22.04 + 0.44 = 37.8

Step 5: Calculate the Standard deviation of X and Y

S.D = sqrt(variance)

S.D(X) = 2.86, S.D(Y) = 3.77

Step6: Calculate the correlation coefficient (r):

r = Σ((Xi - X̄) \* (Yi - Ȳ)) / (sX \* sY \* (n - 1))

r = (37.8) / (2.86 \* 3.77 \* (5 - 1)) = 0.875

the correlation coefficient between advertising expenditure and sales is approximately 0.875. This indicates a strong positive correlation

**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.**

Step1: Calculate mean: mean = 167.85

Step2: Calculate Variance:

Variance = ((160 – 167.85)^2 + (170 - 167.85)^2 + (165 - 167.85)^2 + (155 - 167.85)^2 + (175 - 167.85)^2 +(180 - 167.85)^2 + (170 - 167.85)^2) / 6

= 73.80

Step3: Calculate Standard Deviation:

Sqrt(73.80) = 8.59

**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.**

Steps to be taken to calculate Linear Regression:

Step1: Calculate the mean of Employee Tenure and Job Satisfaction

Step2: Calculate the deviations of each tenure value from the mean (X - X̄) and the deviations of each satisfaction value from the mean (Y - Ȳ)

Step3: Calculate the product of the deviations: (X - X̄)(Y - Ȳ) for each data point.

Step4: Calculate the squared deviations of tenure values from the mean: (X - X̄)^2 for each data point.

Step5: Calculate the estimated slope (β₁) where β₁ = Σ[(X - X̄)(Y - Ȳ)] / Σ[(X - X̄)^2]

Step6: Calculate the estimated intercept (β₀) where β₀ = Ȳ - β₁ \* X̄

Step7: Calculate the Linear Regression using the equation Y = β₀ + β₁ \* X

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | **Employee Tenure (X)** | **Job Satisfaction (Y)** | **(X - X̄)** | **(Y - Ȳ)** | **product of the deviations** | **squared deviations of tenure values** | **the estimated slope (β₁) β₁ = Σ[(X - X̄)(Y - Ȳ)] / Σ[(X - X̄)^2]** | **estimated intercept (β₀)  β₀ = Ȳ - β₁ \* X̄** | **linear regression equation  Y = β₀ + β₁ \* X** |
|  | 2 | 7 | -1.71 | 0.14 | -0.2394 | 2.9241 |  |  | 7 |
|  | 3 | 8 | -0.71 | 1.14 | -0.8094 | 0.5041 |  |  | 7 |
|  | 5 | 6 | 1.29 | -0.86 | -1.1094 | 1.6641 |  |  | 6 |
|  | 4 | 9 | 0.29 | 2.14 | 0.6206 | 0.0841 |  |  | 6 |
|  | 6 | 5 | 2.29 | -1.86 | -4.2594 | 5.2441 |  |  | 5 |
|  | 2 | 7 | -1.71 | 0.14 | -0.2394 | 2.9241 |  |  | 7 |
|  | 4 | 6 | 0.29 | -0.86 | -0.2494 | 0.0841 |  |  | 6 |
| **Sum** | **26** | **48** |  |  | **-6.2858** | **13.4287** | **-0.468087008** | **8.596603** |  |
| **Average** | **3.71** | **6.86** |  |  |  |  |  |  |  |

**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.**

Steps to calculate the ANOVA

Step1: Calculate the mean of 2 groups

Step2: Calculate the Sum of Squares within group. SSW = Σ (Yᵢ - Ȳᵢ)²

Step3: Calculate Degrees of Freedom within Group. DFW = N – K where

N = Number of observations or values

K =Number of Groups

Step4: Calculate mean square within groups MSW = SSW/DFW

Step5: Calculate Sum of squares between groups SSB = Σ (Ȳᵢ - Ȳ)² where

Ȳᵢ = mean of each group

Ȳ = mean of SSW A and SSW B

Step6: Calculate Degrees of Freedom between groups DFB = K – 1

Step7: Calculate Mean square between groups MSB = SSB / DFB

Step8: Calculate F-Statistics F = MSB / MSW

Step9: Calculate critical F value corresponding to the significance level.

Step10: If calculated F-Statistics is greater than critical value then there is a significant difference else no significant difference.

|  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | **A** | **B** | **ssw A Σ (Yᵢ - Ȳᵢ)²** | **ssw B Σ (Yᵢ - Ȳᵢ)²** | **DFW** | **MSW A** | **MSW B** | **SSB(A&B)** | **DFB** | **MSB = SSB / dfB** | **F = MSB / MSW** | **F value** |
|  | 10 | 15 | 4 | 1 | 3 |  |  |  |  |  |  |  |
|  | 12 | 17 | 0 | 1 |  |  |  |  |  |  |  |
|  | 14 | 16 | 4 | 0 |  |  |  |  |  |  |  |
|  | 11 | 14 | 1 | 4 |  |  |  |  |  |  |  |
|  | 13 | 18 | 1 | 4 |  |  |  |  |  |  |  |
| **sum** | 60 | 80 | 10 | 10 |  | 3.33 | 3.33 | 40 | 1 | 40 | 12.01 | 216 |
| **avg** | 12 | 16 |  |  |  |  |  |  |  |  |  |  |

In this case the calculated F statistic is less than critical value and hence no significant difference.

**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.**

Steps to calculate percentiles

Step1: Total number of ratings = n =10

Step2: Sort it in the ascending order and rank it from 1 to n

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

Step3: To find the *p*th percentile multiply the *p*% = (p/100)\*n = (75/100)\*10 = 7.5 rounded to 7.

Step4: Identify the value in that index: in this example it is 8.

**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.**

Step1: Calculate the mean: 10.15

Step2: Define NULL and Alternate Hypothesis:

Null hypothesis (H0): The mean weight is equal to 10 grams.

Alternative hypothesis (Ha): The mean weight differs from 10 grams.

Step3: Calculate the standard deviation.

Variance = 0.059

Standard Deviation = √0.059 = 0.242

Step4: significance level (α) = 0.05

Step5: Calculate the t-value using t = (x̄ - μ) / (s / √n)

t = (10.15 – 10)/(0.242/√6) = 1.518

Step6: Determine Degrees of Freedom = df = n -1 = 6 – 1 = 5

Step7: Use the t-distribution table or a statistical software to find the critical t-value for the given significance level and degrees of freedom.

T value = 2.57

Step8: Compare the calculated t-value to the critical t-value. If the calculated t-value falls within the critical region (i.e., it is larger than the critical t-value for a two-tailed test), then the null hypothesis is rejected, indicating that the mean weight differs significantly from 10 grams. Otherwise, if the calculated t-value falls outside the critical region, the null hypothesis is not rejected, indicating that there is not enough evidence to conclude a significant difference.

Here T table value = 2.57 and calculated t value = 1.51. This falls between -2.57 to 2.57 hence mean weight differs significantly from 10 grams.

**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.**

Step1: Calculate the row totals and column totals for the contingency table.

Step2: Calculate the expected frequencies for each cell of the contingency table

E = (row total \* column total) / grand total

Step3: Calculate the chi-square statistic using the formula:

χ² = Σ [(O - E)² / E]

Step4: Determine the degrees of freedom (df) for the chi-square distribution, which is calculated as:

df = (number of rows - 1) \* (number of columns - 1)

In this case, df = (2 - 1) \* (5 - 1) = 4.

Step5: Use the chi-square distribution table or a statistical software to find the critical chi-square value for the given significance level and degrees of freedom.

Step6: Compare the calculated chi-square statistic to the critical chi-square value. If the calculated chi-square statistic is greater than the critical chi-square value, then there is a significant difference in the click-through rates between the two designs. Otherwise, if the calculated chi-square statistic is smaller than the critical chi-square value, there is no significant difference

|  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  |  |  |  |  |  | **Row Total** | **exp freq1** | **exp freq2** | **exp freq3** | **exp freq4** | **exp freq5** |
| Design A | 100 | 120 | 110 | 90 | 95 | 515 | 96.1 | 109.4 | 106.74 | 98.73 | 104.07 |
| Design B | 80 | 85 | 90 | 95 | 100 | 450 | 83.9 | 95.6 | 93.26 | 86.27 | 90.93 |
| **Col Total** | 180 | 205 | 200 | 185 | 195 | **965** |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |  |  |  |
|  | **chi2-1** | **chi2-2** | **chi2-3** | **chi2-4** | **chi2-5** |  |  |  |  |  |  |
|  | 0.162 | 1.027 | 0.1 | 0.772 | 0.79 | 2.85 |  |  |  |  |  |
|  | 0.185 | 1.175 | 0.114 | 0.883 | 0.905 | 3.26 |  |  |  |  |  |
|  |  |  |  |  | Total Sum | 6.11 |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |  |  |  |
|  |  | Degrees of Freedom = 4 | | | |  |  | p = 0.191083 | |  |  |
|  |  |  |  |  |  |  |  |  |  |  |  |

there is a significant difference in the click-through rates between the two designs

**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.**

Steps to be followed

Step1: Calculate mean = 7.9

Step2: Calculate Standard Deviations = 1.197

Step3: Confidence Level = 0.05

Step4: For a 95% confidence level, the critical value is approximately 1.96.

Step5: Calculate Standard Error = SE = s / √n = 1.197/10 = 0.119

Step6: Calculate the margin of error = Z \* SE = 1.96 \* 0.119 = 0.233

Step7: Calculate the lower and upper bounds = 7.9 – 0.233 to 7.9+0.233 => 7.66 to 8.13

**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.**

<https://github.com/vtech20/ineuron_ppt/blob/main/Assignment_3_Core_modules_PPT_coding.ipynb>

**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.**

<https://github.com/vtech20/ineuron_ppt/blob/main/Assignment_3_Core_modules_PPT_coding.ipynb>

**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.**

Step1: Sort the data in ascending order: [25, 30, 35, 40, 45, 50, 55, 60, 65, 70].

Step2: Calculate the 25th percentile: (25/100)\*10 = 4

Step3: The value at 4th position is 40 = 25th percentile

Step4: Calculate the 75th percentile: (75/100)\*10 = 7.5 = 60

Step5: IQR = 75percentile – 25percentile = 60 – 40 =20

**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.**

<https://github.com/vtech20/ineuron_ppt/blob/main/Assignment_3_Core_modules_PPT_coding.ipynb>

**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.**

<https://github.com/vtech20/ineuron_ppt/blob/main/Assignment_3_Core_modules_PPT_coding.ipynb>

**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.**

1. Calculate the mean of the satisfaction scores:

Mean = (7 + 8 + 9 + 6 + 8 + 7 + 9 + 7 + 8 + 7) / 10 = 7.6

1. Calculate the sum of squared differences from the mean:

Sum of Squared Differences = (7 - 7.6)^2 + (8 - 7.6)^2 + (9 - 7.6)^2 + (6 - 7.6)^2 + (8 - 7.6)^2 + (7 - 7.6)^2 + (9 - 7.6)^2 + (7 - 7.6)^2 + (8 - 7.6)^2 + (7 - 7.6)^2

1. Divide the sum of squared differences by (N-1) to calculate the variance:

Variance = Sum of Squared Differences / (N - 1)

1. Take the square root of the variance to calculate the standard deviation:

Standard Deviation = sqrt(Variance)

1. Divide the standard deviation by the square root of N to calculate the standard error:

Standard Error = Standard Deviation / sqrt(N)

**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.**

<https://github.com/vtech20/ineuron_ppt/blob/main/Assignment_3_Core_modules_PPT_coding.ipynb>