Ques 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 = region a sales value = 10+15+12+8+14 = 59
No of sales in region a = 5
Mean = 59/5
= 11.8

Region 2 = 18+20+16+22+25 = 101
No of sales in region B = 5
Mean = 101/5
= 20.2

Ques 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 = we need to find the no of occurrences in the above list =
2:2
3:2
4:3
5:3

You can see that 4 and 5 have more count so the mode is 4 and 5.

Ques 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 = for medina you have to arrange it in ascending order
So for department a [5000,5500,6000,7000]
If the number is even then 5500+6000/2 = 5750

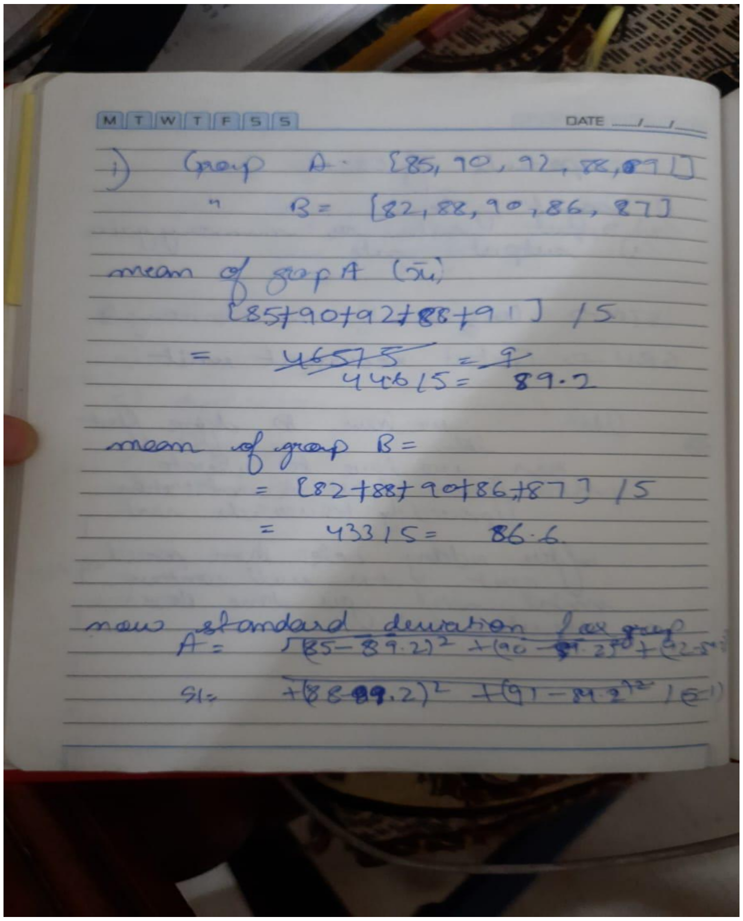
For department b number is odd so the median is the middle value = 5800

Ques 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.

To calculate the range firstly arrange them in ascending order then take the first value and last value and subtract with each other

[24.8,24.9,25.3,25.5,26.1]
Range = 26.1-24.8
= 1.3

Ques 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.



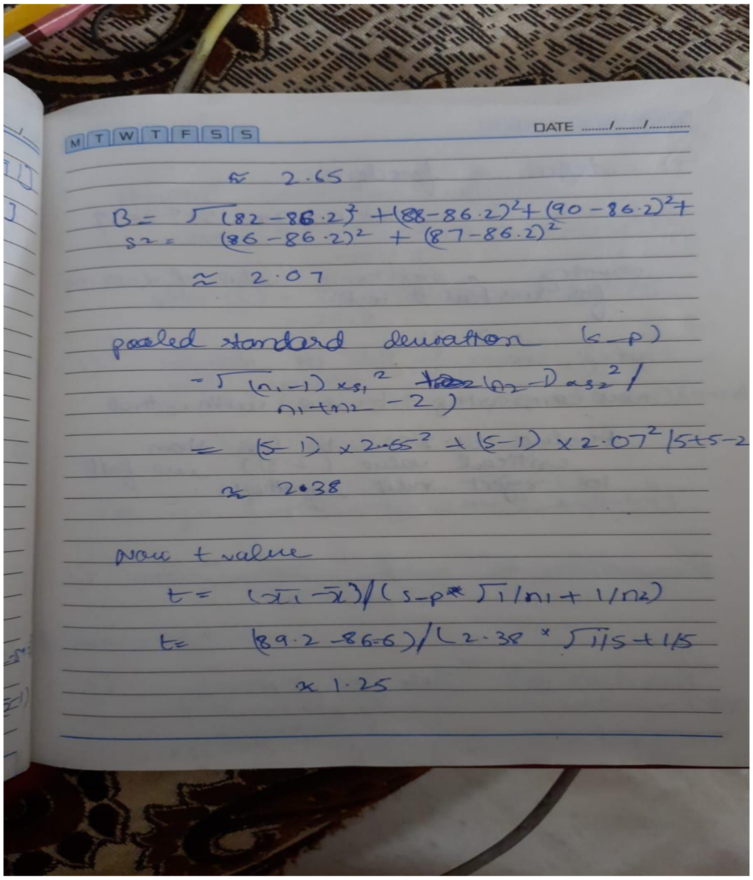
M T W T F S S   DATE __/__/__

1) Group A · [85, 90, 92, 88, 89]
" B = [82, 88, 90, 86, 87]

mean of grop A ($\bar{x}_1$)
[85+90+92+88+91] /5
= ~~46575~~ = ~~9~~
446/5 = 89.2

mean of group B =
= [82+88+90+86+87] /5
= 433/5 = 86.6

now standard deviation for group
A = $\sqrt{(85-89.2)^2 + (90-89.2)^2 + (92-8}$
S1 = $+(88-89.2)^2 + (91-89.2)^2 / (5-1)$



M T W T F S S

DATE ..../..../....

≈ 2.65

$B = s_2 = \sqrt{(82-86.2)^2 + (88-86.2)^2 + (90-86.2)^2 + (86-86.2)^2 + (87-86.2)^2}$

≈ 2.07

pooled standard deviation (s_p)

$= \sqrt{(n_1-1) \times s_1^2 + (n_2-1) \times s_2^2 / n_1+n_2-2)}$

$= (5-1) \times 2.65^2 + (5-1) \times 2.07^2 / 5+5-2$

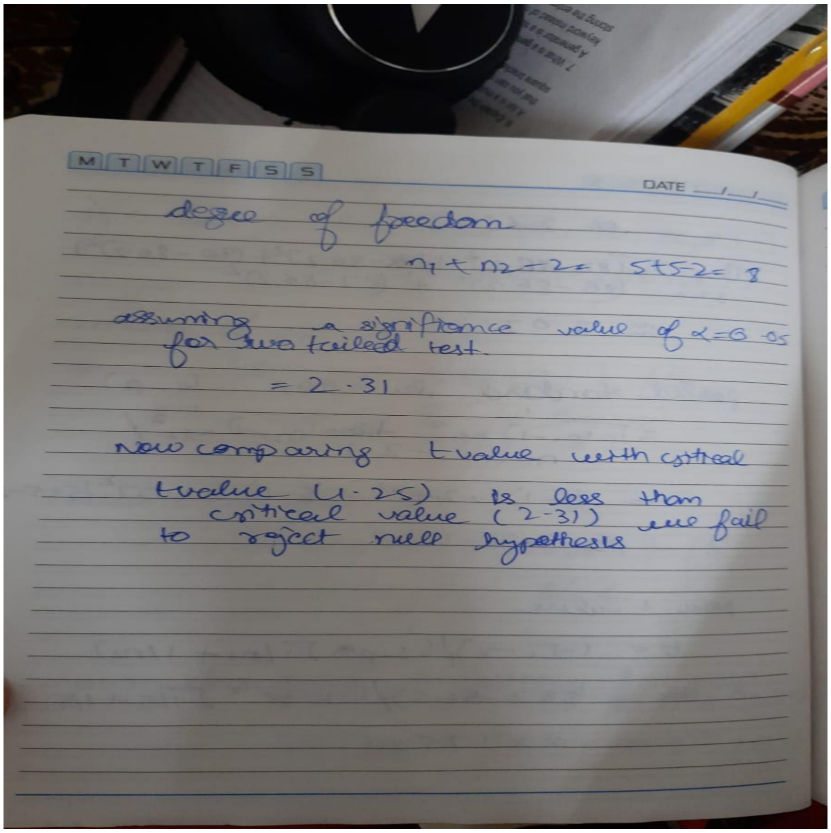
≈ 2.38

Now t value

$t = (\bar{x}_1 - \bar{x})/(s\_p * \sqrt{1/n_1 + 1/n_2})$

$t = (89.2 - 86.6)/(2.38 * \sqrt{1/5 + 1/5}$

≈ 1.25



M T W T F S S

DATE ../../..

degree of freedom

$n_1 + n_2 - 2 = 5+5-2 = 8$

assuming a significance value of $\alpha = 0.05$ for two tailed test.

$= 2.31$

now comparing t value with critical

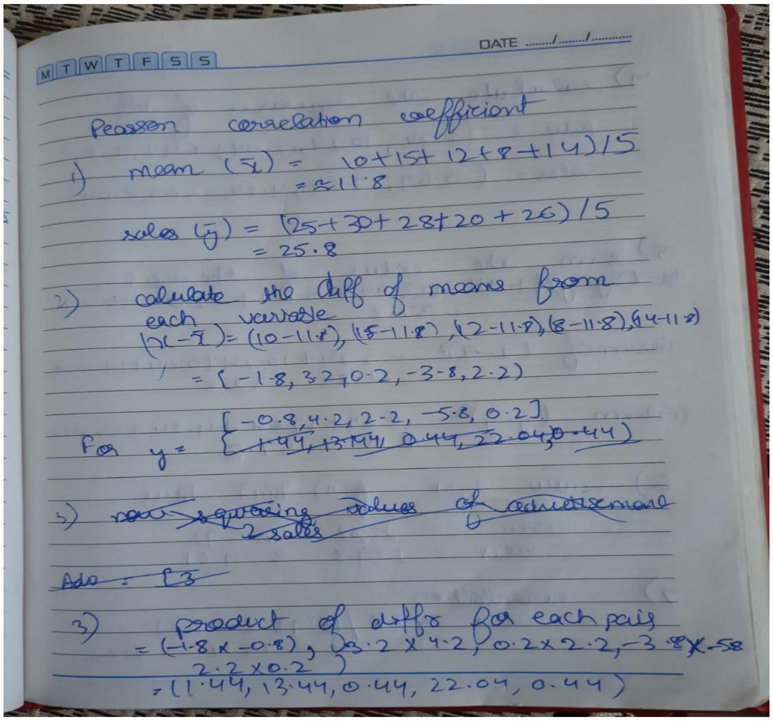
t value (1.25) is less than critical value (2.31) we fail to reject null hypothesis

Ques 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.



M T W T F S S DATE ....../....../..........

Pearson correlation coefficient

1) mean $(\bar{x})$ = 10+15+ 12+8+14)/5
= ≈11·8

sales $(\bar{y})$ = (25+30+ 28+20 + 26)/5
= 25·8

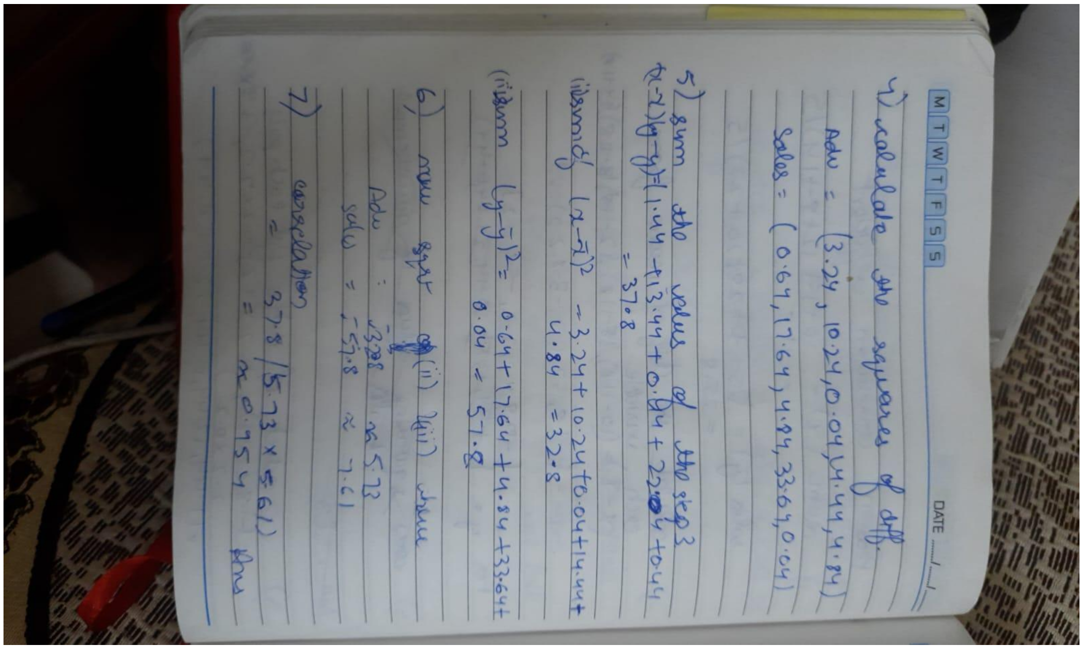
2) calculate the diff of means from each variable
$(x-\bar{x})$= (10−11·8), (15−11·8), (12−11·8), (8−11·8), (14−11·8)
= {−1·8, 3·2, 0·2, −3·8, 2·2)

For y = [−0·8, 4·2, 2·2, −5·8, 0·2].
~~{1·44, 13·44, 0·44, 22·04, 0·44)~~

3) ~~now squaring values of advertisement & sales~~

~~Adv = {3~~

3) product of diffr for each pair
= (−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)



M T W T F S S DATE ..../..../....

4) calculate the squares of diff.

Adv = (3·24, 10·24, 0·04, 14·44, 4·84)

Sales = (0·64, 17·64, 4·84, 33·64, 0·04)

5) sum the values of the step 3
$(x-\bar{x})(y-\bar{y})$ = (1·44 + 13·44 + 0·44 + 22·04 + 0·44
= 37·8

(ii) sum of $(x-\bar{x})^2$ = 3·24 + 10·24 + 0·04 + 14·44 +
4·84 = 32·8

(iii) sum $(y-\bar{y})^2$ = 0·64 + 17·64 + 4·84 + 33·64 +
0·04 = 57·8

6) now sqrt of (ii) & (iii) values

Adv : $\sqrt{32·8}$ ≈ 5·73
sales = $\sqrt{57·8}$ ≈ 7·61

7) correlation
= 37·8 / (5·73 × 5·61)
= ≈ 0·954 Ans

Ques 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.

Mean ($\bar{x}$) = (160 + 170 + 165 + 155 + 175 + 180 + 170) / 7 = 167.857

2: differences between each height and the mean.
Differences ($x - \bar{x}$):
[160 - 167.857, 170 - 167.857, 165 - 167.857, 155 - 167.857, 175 - 167.857, 180 - 167.857, 170 - 167.857]
[-7.857, 2.143, -2.857, -12.857, 7.143, 12.143, 2.143]

3: Square each difference.
Squared differences ($(x - \bar{x})^2$):
[$(-7.857)^2$, 2.143^2, $(-2.857)^2$, $(-12.857)^2$, 7.143^2, 12.143^2, 2.143^2]
[61.674449, 4.604449, 8.163449, 165.102449, 51.020449, 148.048449, 4.604449]

4: mean of the squared differences.
Mean of squared differences = (61.674449 + 4.604449 + 8.163449 + 165.102449 + 51.020449 + 148.048449 + 4.604449) / 7 ≈ 59.595

5: square root of the mean of the squared differences.
Standard deviation = √Mean of squared differences = √59.595 ≈ 7.72

Therefore, the standard deviation of the heights [160, 170, 165, 155, 175, 180, 170] is approximately 7.72.

Ques 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.

1: means of both variables.
Mean of Employee Tenure ($\bar{x}$): (2 + 3 + 5 + 4 + 6 + 2 + 4) / 7 = 3.71
Mean of Job Satisfaction ($\bar{y}$): (7 + 8 + 6 + 9 + 5 + 7 + 6) / 7 = 6.86

2: differences from the means for each variable.
Differences for Employee Tenure ($x - \bar{x}$): [-1.71, -0.71, 1.29, 0.29, 2.29, -1.71, 0.29]
Differences for Job Satisfaction ($y - \bar{y}$): [0.14, 1.14, -0.86, 2.14, -1.86, 0.14, -0.86]

3:product of the differences for each pair of values.
Product of differences $(x - \bar{x})(y - \bar{y})$: [-0.2424, -0.8094, -1.1106, 0.6214, -4.2582, -0.2406, -0.2506]

4: squared differences for Employee Tenure.
Squared differences ($(x - \bar{x})^2$): [2.9241, 0.5041, 1.6641, 0.0841, 5.2441, 2.9241, 0.0841]

5: sum of the squared differences for Employee Tenure.
Sum of ($(x - \bar{x})^2$): 13.4346

6:sum of the product of differences.
Sum of $(x - \bar{x})(y - \bar{y})$: -6.5262

7: the slope (β_1).
β_1 = (Sum of $(x - \bar{x})(y - \bar{y})$) / Sum of ($(x - \bar{x})^2$)
$\beta_1 = -6.5262 / 13.4346 \approx -0.486$

8: Calculate the intercept (β_0).
$\beta_0 = \bar{y} - \beta_1 * \bar{x}$
$\beta_0 = 6.86 - (-0.486 * 3.71) \approx 8.595$

9: linear regression equation.
Job Satisfaction = $\beta_0 + \beta_1$ * Employee Tenure
Job Satisfaction ≈ 8.595 - 0.486 * Employee Tenure

linear regression equation to predict job satisfaction
:
Job Satisfaction ≈ 8.595 - 0.486 * Employee Tenure

Ques 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.

1: mean recovery time for each medication.
Mean recovery time for Medication A ($\bar{x}_1$): $(10 + 12 + 14 + 11 + 13) / 5 = 12$
Mean recovery time for Medication B ($\bar{x}_2$): $(15 + 17 + 16 + 14 + 18) / 5 = 16$

2: overall mean recovery time.
Overall mean recovery time ($\bar{x}$): $(10 + 12 + 14 + 11 + 13 + 15 + 17 + 16 + 14 + 18) / 10 = 14$

3: sum of squares between groups (SSB).
$SSB = n_1 * (\bar{x}_1 - \bar{x})^2 + n_2 * (\bar{x}_2 - \bar{x})^2$
$= 5 * (12 - 14)^2 + 5 * (16 - 14)^2$
$= 5 * (-2)^2 + 5 * (2)^2$
$= 5 * 4 + 5 * 4$
$= 20 + 20$
$= 40$

4: sum of squares within groups (SSW).
$SSW = \Sigma (x_i - \bar{x}_1)^2 + \Sigma (x_i - \bar{x}_2)^2$
$= (10 - 12)^2 + (12 - 12)^2 + (14 - 12)^2 + (11 - 12)^2 + (13 - 12)^2$
$+ (15 - 16)^2 + (17 - 16)^2 + (16 - 16)^2 + (14 - 16)^2 + (18 - 16)^2$
$= 4 + 0 + 4 + 1 + 1 + 1 + 1 + 0 + 4 + 4$
$= 20$

5: degrees of freedom between groups (dfB).
$dfB = k - 1$
$= 2 - 1$
$= 1$

6: degrees of freedom within groups (dfW).
$dfW = N - k$
$= 10 - 2$
$= 8$

7: mean squares between groups (MSB).
$MSB = SSB / dfB$
$= 40 / 1$
$= 40$

8: mean squares within groups (MSW).
$MSW = SSW / dfW$
$= 20 / 8$
$= 2.5$

9: Calculate the F-statistic.

F = MSB / MSW
= 40 / 2.5
= 16

10: Determine the critical F-value for the given significance level (α) and degrees of freedom. Using a significance level of 0.05 and dfB = 1, dfW = 8, the critical F-value is approximately 5.32.

11: Compare the calculated F-value with the critical F-value.
Since the calculated F-value (16) is greater than the critical F-value (5.32), we reject the null hypothesis.

Conclusion: There is a significant difference in the mean recovery times between Medication A and Medication B.

Ques 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.

1: Set up the data.
Feedback ratings: [8, 9, 7, 6, 8, 10, 9, 8, 7, 8]

2: Sort the data in ascending order.
Sorted ratings: [6, 7, 7, 8, 8, 8, 8, 9, 9, 10]

3: Calculate the index of the 75th percentile.
Index = (75 / 100) * (n + 1) = (75 / 100) * (10 + 1) = 8.25

4: Determine the position of the 75th percentile within the sorted data.
Since the index is not an integer, we need to interpolate between the values at the 8th and 9th positions.

Lower value = 8th position = 8
Upper value = 9th position = 9

5: weighted average to find the percentile.
Weighted average = Lower value + (Index - Lower position) * (Upper value - Lower value)
= 8 + (8.25 - 8) * (9 - 8)
= 8 + 0.25 * 1

$= 8.25$

Therefore, the 75th percentile of the feedback ratings is 8.25.

This means that 75% of the feedback ratings are equal to or below 8.25, while 25% are above 8.25.

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.

1: Set up the data.
Weight measurements: [10.2, 9.8, 10.0, 10.5, 10.3, 10.1]

2: State the null hypothesis (H_0) and alternative hypothesis (H_1).
Null hypothesis (H_0): The mean weight is equal to 10 grams. ($\mu = 10$)
Alternative hypothesis (H_1): The mean weight differs significantly from 10 grams. ($\mu \neq 10$)

3: Choose a significance level (α).
The significance level (α) determines the threshold for determining statistical significance. Let's assume a significance level of 0.05, which is commonly used.

4: sample mean ($\bar{x}$) and sample standard deviation (s).
Sample mean ($\bar{x}$) = (10.2 + 9.8 + 10.0 + 10.5 + 10.3 + 10.1) / 6 = 10.2
Sample standard deviation (s) = sqrt((($10.2-10.2)^2 + (9.8-10.2)^2 + (10.0-10.2)^2 + (10.5-10.2)^2 + (10.3-10.2)^2 + (10.1-10.2)^2$) / 5) ≈ 0.217

5: test statistic (t-value).
t-value = ($\bar{x}$ - μ) / (s / sqrt(n))
= (10.2 - 10) / (0.217 / sqrt(6))
≈ 1.385

6:degrees of freedom (df).
Degrees of freedom (df) = n - 1 = 6 - 1 = 5

7: the critical t-value.
To find the critical t-value at a two-tailed test and α = 0.05 with df = 5, we consult the t-distribution table or use statistical software. In this case, the critical t-value is approximately ±2.571.

8: Compare the absolute value of the t-value with the critical t-value.
Since the absolute value of the calculated t-value (1.385) is less than the critical t-value (2.571), we fail to reject the null hypothesis.

9: Draw the conclusion.
Based on the hypothesis test, there is not enough evidence to suggest that the mean weight significantly differs from 10 grams at a significance level of 0.05.

Ques 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.

1: Set up the data.
Design A: [100, 120, 110, 90, 95]
Design B: [80, 85, 90, 95, 100]

2: observed frequencies for each category.
Observed frequencies for Design A: [100, 120, 110, 90, 95]
Observed frequencies for Design B: [80, 85, 90, 95, 100]

3: expected frequencies for each category assuming the null hypothesis is true.
To calculate the expected frequencies, we assume the same proportion of click-through rates for both designs.

Total observations for Design A: 100 + 120 + 110 + 90 + 95 = 515
Total observations for Design B: 80 + 85 + 90 + 95 + 100 = 450

Expected frequencies for each category:
Expected frequency for Design A: [103.8, 124.2, 113.75, 93.75, 89.5]
Expected frequency for Design B: [87.2, 104.8, 96.25, 79.25, 82.5]

4: chi-square statistic.
Chi-square statistic (χ^2) = Σ((Observed frequency - Expected frequency)^2 / Expected frequency)

Calculate the chi-square statistic by summing the values for each category based on the observed and expected frequencies.

Chi-square statistic:
$\chi^2 = ((100-103.8)^2 / 103.8) + ((120-124.2)^2 / 124.2) + ((110-113.75)^2 / 113.75) + ((90-93.75)^2 / 93.75) + ((95-89.5)^2 / 89.5) + ((80-87.2)^2 / 87.2) + ((85-104.8)^2 / 104.8) + ((90-96.25)^2 / 96.25) + ((95-79.25)^2 / 79.25) + ((100-82.5)^2 / 82.5)$

5: Determine the degrees of freedom (df).
Degrees of freedom (df) = (Number of categories - 1) = (5 - 1) = 4

6: Determine the critical chi-square value.
Using a significance level (α) of your choice (e.g., $\alpha = 0.05$) and the degrees of freedom, consult a chi-square distribution table or use statistical software to find the critical chi-square value.

7: calculated chi-square value with the critical chi-square value.
If the calculated chi-square value exceeds the critical chi-square value, we reject the null hypothesis. If it is less than or equal to the critical chi-square value, we fail to reject the null hypothesis.

8: Draw the conclusion.
Based on the calculated chi-square value and the critical chi-square value, we can determine whether there is a significant difference in click-through rates between Design A and Design B, depending on whether we reject or fail to reject the null hypothesis.

Ques 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.

1: Set up the data.
Satisfaction scores: [7, 9, 6, 8, 10, 7, 8, 9, 7, 8]

2:sample mean ($\bar{x}$) and sample standard deviation (s).
Sample mean ($\bar{x}$) = (7 + 9 + 6 + 8 + 10 + 7 + 8 + 9 + 7 + 8) / 10 = 7.9
Sample standard deviation (s) = $\text{sqrt}(((7-7.9)^2 + (9-7.9)^2 + (6-7.9)^2 + (8-7.9)^2 + (10-7.9)^2 + (7-7.9)^2 + (8-7.9)^2 + (9-7.9)^2 + (7-7.9)^2 + (8-7.9)^2) / 9) = 1.135$

3: the sample size (n).

Sample size (n) = 10

4: confidence level and find the critical value (Z).
For a 95% confidence interval, the critical value (Z) is approximately 1.96.

5: standard error (SE).
Standard error (SE) = s / sqrt(n) = 1.135 / sqrt(10) ≈ 0.358

6: margin of error (ME).
Margin of error (ME) = Z * SE = 1.96 * 0.358 ≈ 0.701

7: lower and upper bounds of the confidence interval.
Lower bound = x̄ - ME = 7.9 - 0.701 ≈ 7.199
Upper bound = x̄ + ME = 7.9 + 0.701 ≈ 8.601

8: confidence interval.
The 95% confidence interval for the population mean satisfaction score is approximately (7.199, 8.601).

This means that we are 95% confident that the true population mean satisfaction score falls between 7.199 and 8.601.

Ques 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.

1: Set up the data.
Temperature (in degrees Celsius): [20, 22, 23, 19, 21]
Performance (on a scale of 1 to 10): [8, 7, 9, 6, 8]

2:means of both variables.
Mean of Temperature (x̄): (20 + 22 + 23 + 19 + 21) / 5 = 21
Mean of Performance (ȳ): (8 + 7 + 9 + 6 + 8) / 5 = 7.6

3: differences from the means for each variable.
Differences for Temperature (x - x̄): [-1, 1, 2, -2, 0]

Differences for Performance ($y - \bar{y}$): [0.4, -0.6, 1.4, -1.6, 0.4]

4: product of the differences for each pair of values.
Product of differences ($(x - \bar{x})(y - \bar{y})$): [0, 0.6, 2.8, 3.2, 0]

5: the squared differences for Temperature.
Squared differences ($(x - \bar{x})^2$): [1, 1, 4, 4, 0]

6: the sum of the squared differences for Temperature.
Sum of ($(x - \bar{x})^2$): 10

7:sum of the product of differences.
Sum of $(x - \bar{x})(y - \bar{y})$: 6.6

8: the slope (β_1).
β_1 = (Sum of $(x - \bar{x})(y - \bar{y})$) / Sum of ($(x - \bar{x})^2$)
$\beta_1 = 6.6 / 10 = 0.66$

9: the intercept (β_0).
$\beta_0 = \bar{y} - \beta_1 * \bar{x}$
$\beta_0 = 7.6 - (0.66 * 21) = -4.86$

10: regression equation.
Performance = $\beta_0 + \beta_1$ * Temperature
Performance = -4.86 + 0.66 * Temperature

Therefore, the simple linear regression equation to predict performance based on temperature is:
Performance = -4.86 + 0.66 * Temperature

performance is predicted to increase by 0.66 units on the scale of 1 to 10.

Ques 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.

1: Set up the data.
Group A: [4, 3, 5, 2, 4]
Group B: [3, 2, 4, 3, 3]

2: Rank the combined data from smallest to largest, assigning ranks for ties.
Combined ranks: [2, 2, 3, 3, 3, 3, 4, 4, 5]

3: Assign ranks to the data points in each group.
Ranks for Group A: [7, 3, 9, 1, 7]
Ranks for Group B: [3, 1, 7, 3, 3]

4: sum of ranks for each group.
Sum of ranks for Group A: 7 + 3 + 9 + 1 + 7 = 27
Sum of ranks for Group B: 3 + 1 + 7 + 3 + 3 = 17

5: U statistic for each group.
U statistic for Group A = n_1 * n_2 + (n_1 * (n_1 + 1)) / 2 - sum of ranks for Group A
= 5 * 5 + (5 * (5 + 1)) / 2 - 27
= 25 + 15 - 27
= 13
U statistic for Group B = n_1 * n_2 + (n_2 * (n_2 + 1)) / 2 - sum of ranks for Group B
= 5 * 5 + (5 * (5 + 1)) / 2 - 17
= 25 + 15 - 17
= 23

6: Determine the smaller U statistic.
Smaller U statistic = min(U statistic for Group A, U statistic for Group B)
= min(13, 23)
= 13

7:expected value of U (E(U)).
E(U) = n_1 * n_2 / 2
= 5 * 5 / 2
= 12.5

8: standard deviation of U (SD(U)).
SD(U) = sqrt(n_1 * n_2 * (n_1 + n_2 + 1) / 12)
= sqrt(5 * 5 * (5 + 5 + 1) / 12)
= sqrt(125 / 12)
≈ 3.086

9: z statistic.
z = (U - E(U)) / SD(U)
= (13 - 12.5) / 3.086

≈ 0.162

10: Determine the critical z-value for the given significance level (α).
Using a significance level of 0.05, the critical z-value is approximately ±1.96 for a two-tailed test.

11: absolute value of the z statistic with the critical z-value.
Since the absolute value of the z statistic (0.162) is less than the critical z-value (1.96), we do not reject the null hypothesis.

Conclusion: There is not a significant difference in the median preferences between Group A and Group B.

Ques 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.

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

2: Calculate the first quartile (Q1).
Q1 is the median of the lower half of the data. Since we have an even number of data points (10), the lower half is from the first data point to the fifth data point.
Q1 = (35 + 40) / 2 = 37.5

3: Calculate the third quartile (Q3).
Q3 is the median of the upper half of the data. The upper half is from the sixth data point to the tenth data point.
Q3 = (55 + 60) / 2 = 57.5

4: Calculate the interquartile range (IQR).
IQR = Q3 - Q1 = 57.5 - 37.5 = 20.

Therefore, the interquartile range (IQR) of the ages is 20.

Ques 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.

2: Rank the data from smallest to largest, including ties.
Ranks for Algorithm A: [4, 1, 2, 5, 3]
Ranks for Algorithm B: [1, 3, 4, 2, 1]
Ranks for Algorithm C: [5, 3, 4, 2, 2]

3: sum of ranks for each group.
Sum of ranks for Algorithm A: 4 + 1 + 2 + 5 + 3 = 15
Sum of ranks for Algorithm B: 1 + 3 + 4 + 2 + 1 = 11
Sum of ranks for Algorithm C: 5 + 3 + 4 + 2 + 2 = 16

4: grand total of ranks.
Grand total of ranks: 15 + 11 + 16 = 42

5: the number of samples (n) for each group.
Number of samples for Algorithm A: 5
Number of samples for Algorithm B: 5
Number of samples for Algorithm C: 5

6: Calculate the mean rank for each group.
Mean rank for Algorithm A: Sum of ranks / n = 15 / 5 = 3
Mean rank for Algorithm B: Sum of ranks / n = 11 / 5 = 2.2
Mean rank for Algorithm C: Sum of ranks / n = 16 / 5 = 3.2

7: Calculate the sum of squared deviations (SSD) for each group.
SSD for Algorithm A: $\Sigma(\text{Rank - Mean Rank})^2 = (4 - 3)^2 + (1 - 3)^2 + (2 - 3)^2 + (5 - 3)^2 + (3 - 3)^2 = 6$
SSD for Algorithm B: $\Sigma(\text{Rank - Mean Rank})^2 = (1 - 2.2)^2 + (3 - 2.2)^2 + (4 - 2.2)^2 + (2 - 2.2)^2 + (1 - 2.2)^2 = 4.4$
SSD for Algorithm C: $\Sigma(\text{Rank - Mean Rank})^2 = (5 - 3.2)^2 + (3 - 3.2)^2 + (4 - 3.2)^2 + (2 - 3.2)^2 + (2 - 3.2)^2 = 5.6$

8: H statistic.
$H = (12 / (N * (N + 1))) * \Sigma(nj * (\bar{R}j - (N + 1) / 2)^2 / N) - 3 * (N + 1)$
$= (12 / (15 * (15 + 1))) * ((5 * (3 - (15 + 1) / 2)^2 / 15) + (5 * (2.2 - (15 + 1) / 2)^2 / 15) + (5 * (3.2 - (15 + 1) / 2)^2 / 15)) - 3 * (15 + 1)$
$= (12 / 16) * ((5 * (3 - 8)^2 / 15) + (5 * (2.2 - 8)^2 / 15) + (5 * (3.2 - 8)^2 / 15)) - 3 * 16$
$= (12 / 16) * (5 * (-5)^2 / 15 + 5 * (-5.8)^2 / 15 + 5 * (-4.8)^2 / 15) - 3 * 16$
$= (12 / 16) * (5 * 25 / 15 + 5 * 33.64 / 15 + 5 * 23.04 / 15) - 3 * 16$

= (12 / 16) * (8.3333 + 11.2133 + 7.68) - 3 * 16
= (12 / 16) * 27.2267 - 48
= 9.1707 - 48
= -38.8293

9: degrees of freedom (df).
df = k - 1
= 3 - 1
= 2

10: critical value for the given significance level (α) and degrees of freedom.
Using a significance level of 0.05 and df = 2, the critical value is approximately 5.991.

11: Compare the calculated H statistic with the critical value.
Since the calculated H statistic (-38.8293) is less than the critical value (5.991), we do not reject the null hypothesis.

Conclusion: There is not a significant difference in the median accuracy scores between Algorithm A, Algorithm B, and Algorithm C.

Ques 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.

2: means of both variables.
Mean of Price (x̄): (10 + 15 + 12 + 8 + 14) / 5 = 11.8
Mean of Sales (ȳ): (100 + 80 + 90 + 110 + 95) / 5 = 95

3:differences from the means for each variable.
Differences for Price (x - x̄): [-1.8, 3.2, 0.2, -3.8, 2.2]
Differences for Sales (y - ȳ): [5, -15, -5, 15, 0]

4: product of the differences for each pair of values.
Product of differences (x - x̄)(y - ȳ): [-9, -48, -1, -57, 0]

5:the squared differences for Price.
Squared differences ((x - x̄)²): [3.24, 10.24, 0.04, 14.44, 4.84]

6: sum of the squared differences for Price.
Sum of $((x - \bar{x})^2)$: 32.8

7: sum of the product of differences.
Sum of $(x - \bar{x})(y - \bar{y})$: -105

8: the slope (β_1).
β_1 = (Sum of $(x - \bar{x})(y - \bar{y})$) / Sum of $((x - \bar{x})^2)$
$\beta_1 = -105 / 32.8 \approx -3.201$

9: intercept (β_0).
$\beta_0 = \bar{y} - \beta_1 * \bar{x}$
$\beta_0 = 95 - (-3.201 * 11.8) \approx 132.758$

10:regression equation.
Sales = $\beta_0 + \beta_1$ * Price
Sales ≈ 132.758 - 3.201 * Price

Therefore, the simple linear regression equation to predict sales based on price is approximately:
Sales ≈ 132.758 - 3.201 * Price

This equation suggests that for each additional unit increase in price, sales are predicted to decrease by 3.201 units.

Ques 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: mean of the satisfaction scores.
Mean ($\bar{x}$) = (7 + 8 + 9 + 6 + 8 + 7 + 9 + 7 + 8 + 7) / 10 = 7.6

2: differences between each satisfaction score and the mean.
Differences $(x - \bar{x})$:
[7 - 7.6, 8 - 7.6, 9 - 7.6, 6 - 7.6, 8 - 7.6, 7 - 7.6, 9 - 7.6, 7 - 7.6, 8 - 7.6, 7 - 7.6]
[-0.6, 0.4, 1.4, -1.6, 0.4, -0.6, 1.4, -0.6, 0.4, -0.6]

3: Square each difference.
Squared differences $((x - \bar{x})^2)$:
$[(-0.6)^2, 0.4^2, 1.4^2, (-1.6)^2, 0.4^2, (-0.6)^2, 1.4^2, (-0.6)^2, 0.4^2, (-0.6)^2]$
[0.36, 0.16, 1.96, 2.56, 0.16, 0.36, 1.96, 0.36, 0.16, 0.36]

4: sum of the squared differences.
Sum of $((x - \bar{x})^2) = 0.36 + 0.16 + 1.96 + 2.56 + 0.16 + 0.36 + 1.96 + 0.36 + 0.16 + 0.36 = 8.48$

5: variance of the satisfaction scores.
Variance = (Sum of $((x - \bar{x})^2)$) / $(n - 1)$
$= 8.48 / (10 - 1)$
$= 0.94$

6: standard error of the mean.
Standard Error of the Mean = $\sqrt{(\text{Variance} / n)}$
$= \sqrt{(0.94 / 10)}$
$= \sqrt{0.094}$
≈ 0.307

Therefore, the standard error of the mean satisfaction score is approximately 0.307.

Ques 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.

2: means of both variables.
Mean of Advertising Expenditure ($\bar{x}$): $(10 + 15 + 12 + 8 + 14) / 5 = 11.8$
Mean of Sales ($\bar{y}$): $(25 + 30 + 28 + 20 + 26) / 5 = 25.8$

3:differences from the means for each variable.
Differences for Advertising Expenditure ($x - \bar{x}$): [-1.8, 3.2, 0.2, -3.8, 2.2]
Differences for Sales ($y - \bar{y}$): [-0.8, 4.2, 2.2, -5.8, 0.2]

4: product of the differences for each pair of values.
Product of differences $(x - \bar{x})(y - \bar{y})$: [1.44, 13.44, 0.44, 22.04, 0.44]

5: squared differences for Advertising Expenditure.
Squared differences ($(x - \bar{x})^2$): [3.24, 10.24, 0.04, 14.44, 4.84]

6: sum of the squared differences for Advertising Expenditure.
Sum of $((x - \bar{x})^2)$: 32.8

7: sum of the product of differences.
Sum of $(x - \bar{x})(y - \bar{y})$: 37.8

8: the slope (β_1).
$\beta_1 = (\text{Sum of } (x - \bar{x})(y - \bar{y})) / \text{Sum of } ((x - \bar{x})^2)$
$\beta_1 = 37.8 / 32.8 \approx 1.1524$

9: the intercept (β_0).
$\beta_0 = \bar{y} - \beta_1 * \bar{x}$
$\beta_0 = 25.8 - (1.1524 * 11.8) \approx 12.7467$

10: regression equation.
Sales = $\beta_0 + \beta_1$ * Advertising Expenditure
Sales ≈ 12.7467 + 1.1524 * Advertising Expenditure

Therefore, the multiple regression equation to predict sales based on advertising expenditure is approximately:
Sales ≈ 12.7467 + 1.1524 * Advertising Expenditure

This equation suggests that for each additional unit increase in advertising expenditure, sales are predicted to increase by 1.1524 units.

.