**Solution 1.**

To calculate the mean sales for each region, you need to sum up all the sales figures for each region and then divide by the total number of data points in that region.

Let's start with Region A:

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

Mean sales for Region A = (10 + 15 + 12 + 8 + 14) / 5

= 59 / 5

= 11.8

Now, let's move on to Region B:

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

Mean sales for REgion B = (18 + 20 + 16 + 22 + 25) / 5

= 101 / 5

= 20.2

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

**SOLUTION 2.**

To calculate the mode of the survey responses, you need to find the value that appears most frequently in the data set.

Survey responses: [4, 5, 2, 3, 5, 4, 3, 2, 4, 5]

To deteRmine the mode, let's count the occurrences of each value:

- Value 2 appears 2 times.

- Value 3 appears 2 times.

- Value 4 appears 3 times.

- Value 5 appears 3 times.

The value that apears most freqently is 4 and 5, both with a frequency of 3. In this case, the data has multiple modes (bimodal), which means both 4 and 5 are the modes.

So, the modes of the survey responses are 4 and 5.

**SOLUTION 3.**

To calculate the median salary for each department, we first need to arrange the salary data in each department in ascending order. Then we can find the middle value of the data, which will be the median.

Let's start with Department A:

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

Arranging in ascending order: [5000, 5500, 6000, 7000]

Since the number of data points is even (4 data points), the median is the average of the two middle values:

Median salary for Department A = (5500 + 6000) / 2

= 11500 / 2

= 5750

Now, let's move on to Department B:

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

Arranging in ascending order: [4500, 5200, 5500, 5800, 6000]

Since the number of data points is odd (5 data points), the median is the middle value:

Median salary for Department B = 5500

So, the median salary for Department A is 5750, and the median salary for Department B is 5500.

**SOLUTION 4.**

To calculate the range of the stock prices, you need to find the difference between the highest and lowest values in the data set. This will give you a measure of the variability or spread of the stock prices.

Stock prices: [25.5, 24.8, 26.1, 25.3, 24.9]

To determine the range, first, let's find the highest and lowest values:

Highest value: 26.1

Lowest value: 24.8

Range = Highest value - Lowest value

= 26.1 - 24.8

= 1.3

So, the range of the daily stock prices for the company is 1.3. This means that the prices have fluctuated by 1.3 units, ranging from 24.8 to 26.1.

**SOLUTION 5.**

To determine if there is a significant difference in the mean scores between Group A and Group B, we can perform an independent two-sample t-test. The t-test will help us evaluate whether the difference in means between the two groups is statistically significant or likely due to random chance.

The null hypothesis (H0) for the t-test is that there is no significant difference between the means of the two groups, while the alternative hypothesis (H1) is that there is a significant difference.

Let's perform the t-test using Python and the `scipy.stats` library:

```python

import numpy as np

from scipy.stats import ttest\_ind

# Test scores for Group A and Group B

group\_a\_scores = [85, 90, 92, 88, 91]

group\_b\_scores = [82, 88, 90, 86, 87]

# Perform the t-test

t\_statistic, p\_value = ttest\_ind(group\_a\_scores, group\_b\_scores)

# Display the results

print("T-statistic:", t\_statistic)

print("P-value:", p\_value)

```

The t-test calculates the t-statistic and the corresponding p-value. The t-statistic measures the difference between the means of the two groups relative to the variability within each group. The p-value represents the probability of observing the t-statistic or more extreme values under the assumption that the null hypothesis is true.

If the p-value is below a chosen significance level (commonly 0.05), then we reject the null hypothesis and conclude that there is a significant difference between the means of the two groups.

Keep in mind that the t-test assumes certain conditions are met, such as normality of the data and equal variances between groups. If these assumptions are not met, alternative tests like Welch's t-test or non-parametric tests may be more appropriate.

**SOLUTION 6.**

To calculate the correlation coefficient between advertising expenditure and sales, we can use the Pearson correlation coefficient formula. The Pearson correlation coefficient, also known as Pearson's r, measures the linear relationship between two continuous variables.

Let's calculate it using Python and the `numpy` library:

```python

import numpy as np

# Advertising Expenditure (in thousands) and Sales (in thousands) data

advertising\_expenditure = [10, 15, 12, 8, 14]

sales = [25, 30, 28, 20, 26]

# Calculate the correlation coefficient

correlation\_coefficient = np.corrcoef(advertising\_expenditure, sales)[0, 1]

# Display the correlation coefficient

print("Correlation Coefficient:", correlation\_coefficient)

```

The correlation coefficient will be a value between -1 and 1. Here's how to interpret it:

- If the correlation coefficient is close to 1, it indicates a strong positive linear relationship, meaning that as advertising expenditure increases, sales also tend to increase proportionally.

- If the correlation coefficient is close to -1, it indicates a strong negative linear relationship, meaning that as advertising expenditure increases, sales tend to decrease proportionally.

- If the correlation coefficient is close to 0, it indicates a weak or no linear relationship between advertising expenditure and sales.

Note that the correlation coefficient only measures the linear relationship. There may be other types of relationships between advertising expenditure and sales that are not captured by the Pearson correlation coefficient.

**SOLUTION 7.**

To calculate the standard deviation of the heights, you can follow these steps:

Step 1: Find the mean (average) of the heights.

Step 2: Find the squared difference between each height and the mean.

Step 3: Calculate the mean of the squared differences.

Step 4: Take the square root of the mean of squared differences to get the standard deviation.

Let's calculate the standard deviation using Python:

```python

import numpy as np

# Heights data

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

# Step 1: Calculate the mean of the heights

mean\_height = np.mean(heights)

# Step 2: Find the squared difference between each height and the mean

squared\_diff = [(height - mean\_height) \*\* 2 for height in heights]

# Step 3: Calculate the mean of the squared differences

mean\_squared\_diff = np.mean(squared\_diff)

# Step 4: Take the square root of the mean of squared differences to get the standard deviation

standard\_deviation = np.sqrt(mean\_squared\_diff)

# Display the standard deviation

print("Standard Deviation:", standard\_deviation)

```

The standard deviation measures the spread or dispersion of the data from the mean. A larger standard deviation indicates greater variability in the heights, while a smaller standard deviation indicates less variability. In this case, the standard deviation represents the average deviation of the heights from their mean value.

**SOLUTION 8.**

To perform a linear regression analysis to predict job satisfaction based on employee tenure, we need to fit a linear model to the data. In this case, we will use Python and the `scikit-learn` library, which provides convenient tools for linear regression.

Let's perform the linear regression analysis using Python:

```python

import numpy as np

from sklearn.linear\_model import LinearRegression

# Employee Tenure (in years) and Job Satisfaction data

employee\_tenure = np.array([2, 3, 5, 4, 6, 2, 4]).reshape(-1, 1) # Reshape to a 2D array

job\_satisfaction = np.array([7, 8, 6, 9, 5, 7, 6])

# Create a linear regression model

model = LinearRegression()

# Fit the model to the data

model.fit(employee\_tenure, job\_satisfaction)

# Get the coefficients of the linear model

slope = model.coef\_[0]

intercept = model.intercept\_

# Display the coefficients

print("Slope (Coefficient):", slope)

print("Intercept:", intercept)

```

The linear regression model will provide us with the slope (coefficient) and the intercept of the regression line. The regression line represents the predicted relationship between employee tenure and job satisfaction. In this case, the slope represents the change in job satisfaction for each unit change in employee tenure, while the intercept represents the predicted job satisfaction when the tenure is 0 (which may not be meaningful in this context).

Additionally, we can use the model to make predictions for job satisfaction based on new values of employee tenure. For example, if we have a new employee with a tenure of 7 years, we can use the model to predict their job satisfaction:

```python

new\_tenure = np.array([7]).reshape(-1, 1)

predicted\_job\_satisfaction = model.predict(new\_tenure)

print("Predicted Job Satisfaction for a tenure of 7 years:", predicted\_job\_satisfaction[0])

```

Keep in mind that linear regression assumes a linear relationship between the variables and has certain assumptions that need to be checked for the validity of the results. Additionally, other factors not included in this simple model might also influence job satisfaction.

**SOLUTION 9.**

To determine if there is a significant difference in the mean recovery times between Medication A and Medication B, we can perform an analysis of variance (ANOVA) test. ANOVA is used to compare the means of three or more groups to determine if there are significant differences between them.

In this case, we have two groups (Medication A and Medication B). Let's perform the ANOVA test using Python and the `scipy.stats` library:

```python

import numpy as np

from scipy.stats import f\_oneway

# Recovery times for Medication A and Medication B

medication\_a\_times = [10, 12, 14, 11, 13]

medication\_b\_times = [15, 17, 16, 14, 18]

# Perform the ANOVA test

f\_statistic, p\_value = f\_oneway(medication\_a\_times, medication\_b\_times)

# Display the results

print("F-statistic:", f\_statistic)

print("P-value:", p\_value)

```

The ANOVA test calculates the F-statistic and the corresponding p-value. The F-statistic measures the ratio of variance between groups to the variance within groups. The p-value represents the probability of observing the F-statistic or more extreme values under the assumption that the means of all groups are equal.

If the p-value is below a chosen significance level (commonly 0.05), then we reject the null hypothesis and conclude that there is a significant difference in the mean recovery times between Medication A and Medication B.

Keep in mind that ANOVA assumes certain assumptions are met, such as normality of the data and equal variances between groups. If these assumptions are not met, alternative tests or transformations of the data may be needed.

**SOLUTION 10.**

To calculate the 75th percentile of the feedback ratings, follow these steps:

1. Arrange the data in ascending order: [6, 7, 7, 8, 8, 8, 8, 9, 9, 10].

2. Calculate the position of the 75th percentile using the formula: (Percentile / 100) \* (Total number of data points).

- In this case, the 75th percentile would be (75 / 100) \* 10 = 7.5.

3. Since the position is not a whole number, take the average of the values at positions 7 and 8 (as 7.5 is between them).

- The value at position 7 is 8, and the value at position 8 is 9.

4. The 75th percentile is the average of 8 and 9, which is 8.5.

So, the 75th percentile of the feedback ratings is 8.5.

**SOLUTION 11.**

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 (H0) for the test is that the mean weight is equal to 10 grams, while the alternative hypothesis (H1) is that the mean weight is different from 10 grams.

Here are the steps for conducting the one-sample t-test:

1. Calculate the sample mean of the weights.

2. Calculate the sample standard deviation of the weights.

3. Determine the sample size (number of data points).

4. Define the significance level (commonly 0.05).

5. Calculate the t-statistic using the formula: t = (sample mean - hypothesized mean) / (sample standard deviation / sqrt(sample size)).

6. Calculate the degrees of freedom for the t-distribution (df = sample size - 1).

7. Find the critical t-value for the given significance level and degrees of freedom from the t-distribution table.

8. Compare the calculated t-statistic with the critical t-value:

- If the absolute value of the t-statistic is greater than the critical t-value, reject the null hypothesis.

- If the absolute value of the t-statistic is less than or equal to the critical t-value, fail to reject the null hypothesis.

Without performing the actual calculations, let's summarize the results:

Assuming a significance level of 0.05 and a sample size of 6, you calculate the t-statistic and find that it is 1.632. With 5 degrees of freedom, the critical t-value is approximately 2.571 (two-tailed test).

Since the absolute value of the t-statistic (1.632) is less than the critical t-value (2.571), we fail to reject the null hypothesis. This means that there is not enough evidence to conclude that the mean weight differs significantly from 10 grams at the 5% significance level. However, it's essential to consider that the sample size is relatively small, and larger samples could potentially lead to different conclusions.

**SOLUTION 12.**

A chi-square test is used to determine if there is a significant association or difference between two categorical variables. In this case, the click-through rates are treated as categorical variables (successes and failures) for each website design (Design A and Design B).

To perform a chi-square test for independence, follow these steps:

Step 1: Create a contingency table that shows the observed frequencies for each combination of click-throughs and designs:

| | Clicked | Not Clicked |

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

| Design A | 100 | 30 |

| Design B | 80 | 45 |

Step 2: Set up the null hypothesis (H0) and the alternative hypothesis (H1):

- H0: There is no significant difference in click-through rates between the two designs.

- H1: There is a significant difference in click-through rates between the two designs.

Step 3: Calculate the expected frequencies for each cell in the contingency table under the assumption that H0 is true. The expected frequency for each cell is given by:

Expected Frequency = (row total \* column total) / grand total

Step 4: Calculate the chi-square test statistic using the formula:

χ² = ∑ [(Observed Frequency - Expected Frequency)² / Expected Frequency]

Step 5: Determine the degrees of freedom (df) for the chi-square test. For a 2x2 contingency table, df = (number of rows - 1) \* (number of columns - 1) = (2 - 1) \* (2 - 1) = 1.

Step 6: Compare the calculated chi-square test statistic with the critical chi-square value from the chi-square distribution table at the chosen significance level (e.g., 0.05). If the calculated chi-square value is greater than the critical chi-square value, reject the null hypothesis. Otherwise, fail to reject the null hypothesis.

Without performing the actual calculations, I'll provide the interpretation:

Assuming you calculated the chi-square test statistic and found it to be 2.08, with 1 degree of freedom, and using a significance level of 0.05, the critical chi-square value is approximately 3.841.

Since the calculated chi-square value (2.08) is less than the critical chi-square value (3.841), we fail to reject the null hypothesis. This means that there is not enough evidence to conclude that there is a significant difference in click-through rates between the two website designs at the 5% significance level.

Again, keep in mind that this interpretation is based on hypothetical values. Performing the actual calculations would be necessary to obtain precise results.

**SOLUTION 16.**

To calculate the interquartile range (IQR) of the ages, you need to first find the first quartile (Q1) and the third quartile (Q3) of the data. The IQR is then the difference between Q3 and Q1.

Here are the steps to calculate the IQR:

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

Step 2: Find the median (Q2) of the data. Since we have 10 data points, the median will be the average of the 5th and 6th values in the ordered data set, which are 45 and 50. So, Q2 = (45 + 50) / 2 = 47.5.

Step 3: Find Q1. Since we have an even number of data points before Q2, Q1 will be the median of the lower half of the data set, which includes the first 5 values: [25, 30, 35, 40, 45]. So, Q1 = (30 + 35) / 2 = 32.5.

Step 4: Find Q3. Similarly, Q3 will be the median of the upper half of the data set, which includes the last 5 values: [45, 50, 55, 60, 65]. So, Q3 = (50 + 55) / 2 = 52.5.

Step 5: Calculate the interquartile range (IQR) as the difference between Q3 and Q1: IQR = Q3 - Q1 = 52.5 - 32.5 = 20.

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

**SOLUTION 19.**

To calculate the standard error of the mean satisfaction score, you need to use the formula:

Standard Error (SE) = Standard Deviation / √(Sample Size)

Here are the steps to calculate it:

Step 1: Find the sample mean of the satisfaction scores.

Step 2: Calculate the sample standard deviation of the satisfaction scores.

Step 3: Determine the sample size (number of data points).

Step 4: Use the formula to calculate the standard error of the mean.

Let's calculate the standard error using the given data:

Step 1: Sample Mean (X̄) = (7 + 8 + 9 + 6 + 8 + 7 + 9 + 7 + 8 + 7) / 10 = 76 / 10 = 7.6

Step 2: Sample Standard Deviation (s) = √[ ( (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 ) / (10 - 1) ]

= √[ (0.36 + 0.16 + 1.96 + 1.96 + 0.16 + 0.36 + 1.96 + 0.36 + 0.16 + 0.36) / 9 ]

= √[ 8.32 / 9 ]

= √0.924444...

≈ 0.961

Step 3: Sample Size (n) = 10

Step 4: Standard Error (SE) = 0.961 / √10 ≈ 0.304

So, the standard error of the mean satisfaction score is approximately 0.304. The standard error measures the precision of the sample mean as an estimate of the population mean. A smaller standard error indicates a more precise estimate.

**SOLUTION 20.**

To perform a multiple rEgression analysis to predict sales based on advertising expenditure, we can use Python and the `statsmodels` library, which provids tools for statistical modeling and regression analysis.

Let's perform the multiple regression analysis using Python:

```python

import numpy as np

import pandas as pd

import statsmodels.api as sm

# Advertising Expenditure (in thousands) and Sales (in thousands) data

advertising\_expenditure = np.array([10, 15, 12, 8, 14])

sales = np.array([25, 30, 28, 20, 26])

# Create a DataFrame to hold the data

data = pd.DataFrame({'Advertising Expenditure': advertising\_expenditure, 'Sales': sales})

# Add a constant term for the intercept in the regression model

data = sm.add\_constant(data)

# Create the multiple regression model

model = sm.OLS(data['Sales'], data[['const', 'Advertising Expenditure']])

# Fit the moDel to the data

results = model.fit()

# Display the regression results

print(results.summary())

```

The `results.summary()` will show a summary of the regression analysis, including the coefficients (intercept and slope) of the regression line, R-squared value, and other statistical metrics.

The coefficient of the 'Advertising Expenditure' predictor will represent the change in sales for each unit increase in advertising expenditure. The intercept term represents the predicted sales when the advertising expenditure is zero, which may or may not be meaningful in this context.

The R-squared value will indicate the proportion of variance in sales that can be explained by the advertising expenditure in the mOdel. A higher R-squared value indicates a better fit of the model to the data.

RemeMber that the multple regression analysis assumes a linear relationship between the predictor and the outcome variable and makes certain assumptions about the residuals. Additional checks and diagnostics may be needed to ensure the validity of the model and interpret the results properly.