PSET 8: T-TEST and ANOVA Exercise.

Analyzing Customer Feedback and Sales Data of Mobile Apps

Objective: Perform a statistical analysis of customer feedback and sales data for a mobile app company using T-Tests and ANOVA.

Data Sources:

1. **Customer Feedback Data** (**customer_feedback.csv**): Contains customer ratings (1-10 scale) for both iOS and Android apps.

2. **Sales Data** (**sales_data.csv**): Monthly sales figures for the same apps over a corresponding period.

Tasks:

1. **Data Preparation**:

   ● Load data from **customer_feedback.csv** and **sales_data.csv** into pandas data frames.

   ● Convert the 'date' column in both datasets to pandas datetime objects for analysis.

2. **Customer Feedback Analysis (feedback_analysis function)**:

   ● Determine the appropriate T-Test (independent, paired, one-tail, or two-tail) based on data and hypothesis.

   ● Convert the feedback scores into a 2-dimensional numpy array (iOS scores as the first dimension, Android scores as the second).

   ● Compute and return the t-test statistic and the pvalue and print the returned values.

   ● Analyze if there's a significant difference in average customer satisfaction between iOS and Android apps.

   ● Interpret the result, i.e., if the p-value is significant for average customer satisfaction between the two groups.

3. **Sales Performance Analysis (sales_analysis function)**:

   ● Compare sales before and after a major marketing campaign (March 1-31, 2023).

   ● Use an appropriate T-Test to assess the campaign's impact on sales.

   ● Return the t-test statistic and the p-value and print the returned values.

   ● Interpret the result, i.e., if a significant impact is found based on the p-value;

4. **Seasonal Sales Analysis (seasonal_analysis function)**:

   ● Examine sales differences between summer (June-August) and winter (December-February).

   ● Apply a T-Test to assess if these variations are statistically significant.

- ● Return the t-test statistic and the p-value and print the returned values.

- ● Interpret the result, i.e., if significant seasonal variations exists based on the p-value.

5. **Feedback Consistency Analysis (consistency_analysis function)**:

- ● Assess if monthly feedback scores are consistent across January, May, September, and December.

- ● Use one-way ANOVA to test for significant differences in feedback scores across these months.

- ● Return the statistic and the p-value and print the returned values.

- ● Interpret the result, i.e., if the difference are significant based on p-value.

6. **Sales and Feedback Correlation Analysis (corr_analysis function)**:

- ● Investigate if high customer feedback correlates with increased sales.

- ● Merge feedback and sales data, categorizing sales into high and low feedback scores.

- ● Perform a T-Test to compare sales in months with high vs. low feedback scores.

- ● Return the statistic and the p-value and print the returned values.

- ● Interpret the result, i.e., if correlation is significant based on the p-value.

Additional Guidance:

- ● Choose the T-Test type based on data characteristics and research questions.

- ● Ensure assumptions like equal variance are correctly handled, especially in independent T-Tests.

- ● Use 0.05 as the significance level for p-value.

- ● For interpreting your results , create a new markdown cell in your jupyter notebook, and write your interpretation in complete sentences.

- ● For question 2-6 we are asking you to filter the data into either 2 or more groups based on some condition and perform statistical tests. Simply filter the data into 2 or more groups. Do not use pandas groupby operation which aggregates the data. For example in question 2 , you will perform a t-test on two groups. The first group contains all feedback scores for android users and the second group contains feedback scores for ios users without any aggregation.

Example Function Headers:

```
def feedback_analysis(df_feedback):

        return statistic, p_val
```

```python
def sales_analysis(df_sales):

        return statistic, p_val


def seasonal_analysis(df_sales):

        return statistic, p_val


def consistency_analysis(df_feedback):

        return statistic, p_val


def corr_analysis(df_feedback, df_sales):
        return statistic, p_val
```