

SUMMARY 3

1. Data Loading:

- Imports the Pandas library and reads a CSV file named "Premotorbike.csv" into a Pandas DataFrame called `dataset`.
- Displays the contents of the `dataset` DataFrame.

2. Handling Missing Values:

- Calculates and displays the count of missing values for each column in the dataset using `dataset.isnull().sum()`.

3. Covariance and Correlation Analysis:

- Calculates the covariance matrix of the dataset using `dataset.cov()`.
- Calculates the correlation matrix of the dataset using `dataset.corr()`.

4. Hypothesis Testing:

- Performs a two-sample paired t-test between the "price" and "mileage" columns for motorcycles with "Manual" gear using `ttest_rel()`. The result is that the null hypothesis is accepted, indicating no significant difference between price and mileage.
- Performs a two-sample paired t-test between the "price" and "power" columns for motorcycles with "Manual" gear using `ttest_rel()`.
- Performs a two-sample independent t-test between the "price" columns for motorcycles with "Manual" and "Automatic" gears using `ttest_ind()`.
- Performs a two-sample independent t-test between the "mileage" columns for motorcycles with "Manual" and "Automatic" gears using `ttest_ind()`.
- Conducts a one-way ANOVA test among the "price," "mileage," and "power" columns using `stats.f_oneway()`. The result suggests no significant difference between the columns.

5. Data Visualization:

- Uses Seaborn to create a distribution plot (`sns.distplot()`) for the "price" column, visualizing the distribution of prices in the dataset.

In summary, this code explores the dataset by calculating statistics, performing hypothesis tests, and visualizing the distribution of prices. It includes tests for covariance, correlation, t-tests, and ANOVA to examine relationships and differences among various columns in the dataset.

SUMMARY 3