

Data Statistics, Mining and Application quiz

TOTAL POINTS 8

1. When doing qualitative data distribution analysis, statistics are often made on the category of an attribute. In addition to the pie chart, the object's `value_counts()` method is often used for calculation. The calculation results are shown in the figure below. Assume that the attribute target has 3 categories, and each category has 50, 20, and 50 samples.

1 point

```
2    50
1    20
0    50
Name: target, dtype: int64
```

Enter answer here

2. Statistical analysis is divided into two categories: central tendency analysis and dispersion tendency analysis. What are the indicators for dispersion tendency analysis in the following options?

1 point

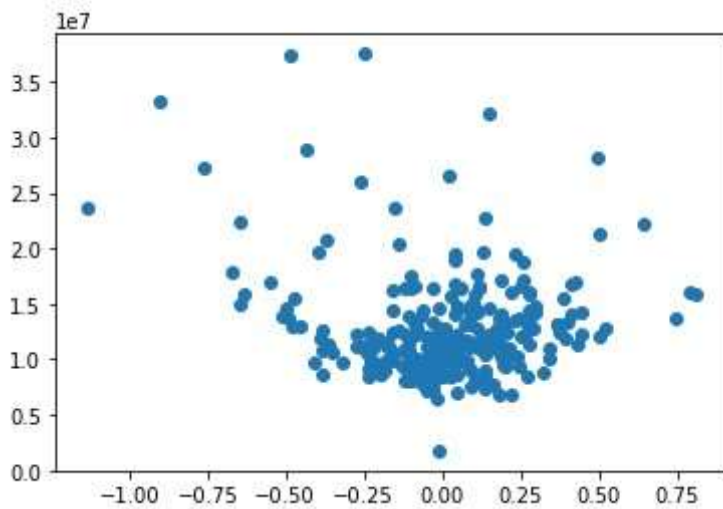
- ☐ Mean
- ☒ Standard deviation
- ☐ Median
- ☒ IRQ

3. Scatter plots are an effective way to observe the relationship between two one-dimensional data sequences. Please fill in the spaces (the two answers are separated by a semicolon) to get the scatter plot of the difference between the daily opening price and the closing price and the current day's trading volume of the Coca-Cola

1 point

company in the past year. Assuming that the data has been obtained and stored in a DataFrame object named `quotesK0df`, the names of the daily closing price, opening price and trading volume are `open`, `close`, and `volume`.

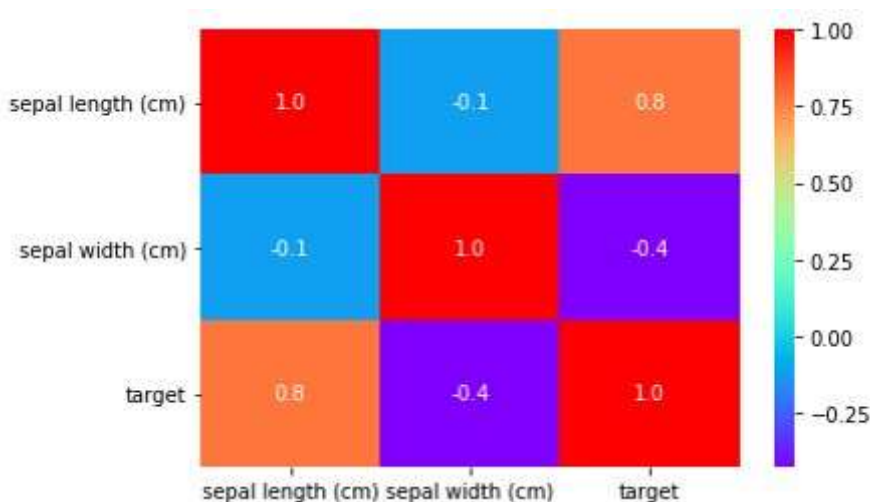
```
1 plt._____(quotesK0df.close - quotesK0df.open, quotesK0df._____)
```



Enter answer here

4. Observe the following heat map based on the Pearson correlation coefficient to determine the correlation between the attributes sepal width (cm) and sepal length (cm). Which of the following options may be met?

1 point



- ☐ Positive strong linear correlation
- ☐ Negative strong linear correlation
- ☐ Positive weak linear correlation
- ☒ Negative weak linear correlation

5. Suppose you want to calculate the median of all attributes that is in the records with opening price in the range [52,54] of the Coca-Cola Company in the past year (assuming that the data is stored in the DataFrame object named quotesK0df). Please choose the options that match the appropriate code at two horizontal lines.

1 point

```
1 quotesK0df[(quotesK0df.open >= 52) _____ (quotesK0df.open <= 54)]
   ._____()
```

- ☐ |; mean
- ☐ or; median
- ☒ &; median
- ☐ and; median

6. Suppose you have obtained some historical data of Coca-Cola Company and stored the data in a DataFrame object named quotesK0df. The date is set to the index of quotesK0df. Which of the following options can be used to calculate the total volume of each month in quotesK0df?

1 point

	close	high	low	open	volume
2018-10-22	45.970001	46.459999	45.810001	46.369999	14283600
2018-10-23	46.360001	46.459999	45.450001	45.619999	17483900
2018-10-24	46.730000	47.389999	46.240002	46.330002	21626000
2018-10-25	46.509998	46.740002	46.119999	46.650002	14220200
2018-10-26	45.919998	46.480000	45.430000	46.369999	21458000



```
1 >>> month = [item[5:7] for item in quotesK0df.index]
2 >>> sum(quotesK0df.groupby(month).volume)
```



```
1 >>> month = [item[5:7] for item in quotesK0df.index]
2 >>> quotesK0df.groupby(month).volume.sum()
```



```
1 >>> month = [item[5:7] for item in quotesK0df.index]
2 >>> quotesK0df.groupby(month).volume.apply(sum)
```



```
1 >>> month = [item[5:7] for item in quotesK0df.index]
2 >>> quotesK0df.groupby(month).volume.apply(sum())
```

7. If you have selected two parts of data from a DataFrame object named df (keep the data attributes complete) and stored them in the DataFrame objects named df1 and df2 respectively, the code is as follows. Please select the function/method from the following options that can correctly merge the two parts of data into one.

1 point

```
1 >>> import pandas as pd
2 >>> pd._____([df1, df2])
```

☐ append

☒ concat

- ☐ join
- ☐ merge

8. Please decide whether the following statements are true or false.

1 point

K-means is an unsupervised clustering learning algorithm; K represents the center quantity initially selected by the user in space.

- ☒ T
- ☐ F

☐ I, **Soujatya Bhattacharya**, understand that submitting work that isn't my own may result in permanent failure of this course or deactivation of my Coursera account.

[Learn more about Coursera's Honor Code](#)