**Question 1 of 6**

You have loaded some data into a DataFrame, as shown below. How should you now create a scatter plot of the data with 'Name' for the x-axis and 'Age' for the y-axis?

data={'Name':['Jon', 'Sara', 'Sakura', 'Yo'], 'Age':[20, 18, 27, 50]} DataFrame = pd.DataFrame(data = data);

* 

DataFrame.plot.scatter(Data\_x = 'Age', Data\_y = 'Name');

* 

**DataFrame.plot.scatter(x = 'Name', y = 'Age');**

**In addition to setting your x and y axis, you can also set the size and color for your scatter plot**.

* 

plotscatter.DataFrame(x = 'Name', y = 'Age');

## Question 2 of 6

You created a scatter plot using the command DataFrame.plot.scatter(x = 'Name', y = 'Age'). How can this be changed so the dots have a size of 100 and color red?

* 

scatter\_plot.DataFrame(x = 'Name', y = 'Age', s = 100, c = 'red');

* 

**DataFrame.plot.scatter(x = 'Name', y = 'Age', s = 100, c = 'red');**

**DataFrame.plot.scatter(x = 'Name', y = 'Age', s = 100, c = 'red');**

**Correct**

**You could also have "DataFrame.plot.scatter('Name', 'Age', 100, 'red');", but it is less clear.**

* DataFrame.plot.scatter('red',100,'Age','Name');

## Question 3 of 6

What should you use to set text to the horizontal axis for a 2D graph?

* **xlabel()**

**xlabel()**

**Correct**

**An example of this could be "DataFrame.xlabel('x-label')".**

* ylabel()
* axis()

## Question 4 of 6

How do you calculate the mean of a data set?

* **Sum the values of the data set, and then divide that sum by the number of values.**

**Sum the values of the data set, and then divide that sum by the number of values.**

**Correct**

**In pandas, use the "mean()" function to easily determine the mean of a DataFrame.**

* Take the sum of each value minus the mean, and then divide by the mean.
* Sum the square of each value of the data set, and then divide by the number of values.

## Question 5 of 6

How do you calculate the variance for a data set?

* Take the square root of the sum of the difference between each value, and then divide by the mean.
* Sum the square of each value of the data set, and then divide by the number of values.
* **Sum the squares of each value minus the mean, and then divide that by the number of values.**

**Sum the squares of each value minus the mean, and then divide that by the number of values.**

**Correct**

**From here, you can determine the standard deviation by taking the square root of the variance.**

## Question 6 of 6

You have loaded a set of data into a DataFrame, as shown below. How can you plot a scatter plot and a line plot on the same plot?

data={'Name':['Jon', 'Sara', 'Sakura', 'Yo'], 'Age':[20, 18, 27, 50]} DataFrame = pd.DataFrame(data = data);

* DataFrame.plot.scatter(x = 'Name', y = 'Age', s = 100, c = 'red'); DataFrame.plot.line(x = 'Name', y = 'Age')
* **ax = DataFrame.plot.scatter(x = 'Name', y = 'Age', s = 100, c = 'red'); DataFrame.plot.line(x = 'Name', y = 'Age',ax=ax)**

**ax = DataFrame.plot.scatter(x = 'Name', y = 'Age', s = 100, c = 'red'); DataFrame.plot.line(x = 'Name', y = 'Age',ax=ax)**

**Correct**

**You have to pass the scatter plot to the line plot in order to have them both on the same plot.**

****DataFrame.plot.line(x = 'Name', y = 'Age') DataFrame.plot.scatter(x = 'Name', y = 'Age', s = 100, c = 'red');

## Question 1 of 2

Which function should you use to display an imported image?

* scatter()
* figure()
* **imshow()**

## Question 2 of 2

What does a 90% confidence interval mean?

* **You are 90% confident that the true value of the population parameter will be contained within the range for a series of runs.**
* **Correct**
* **So, on average, 10 out of 100 runs will have population parameters not contained in the confidence interval.**
* There is a 90% probability that the true population parameter is contained with the confidence interval range.
* You are 90% confident that you collected your data correctly and it is accurate for a large sample of runs.

## Question 1 of 6

How should you describe k-fold cross-validation?

* compiling the data for multiple runs, and then comparing it to a model created heuristically from previous trials
* dividing the data into two parts, and then using two trial runs to train the model
* **dividing the data into k subsets, and then using k successive trials to train and test**
* **Correct**
* **This functionality is supported by "sklearn".**

How should you fit a model to the observed data?

* Alter the predicted response variables until there is no difference between them and the observed.
* Manipulate the observed data statically until the difference between the observed and predicted is minimized.
* **Modify the model parameters until the difference between the predicted and observed data is minimized.**
* **Correct**
* **This difference is generally referred to as the residual.**

In cross-validation, a data set is split into two parts. Which part is used to feed to the model?

* the predication error data
* **the training data**
* **Correct**
* **This is for generating a model using machine learning, and it will be compared to the testing data.**
* the testing data

How is Bayesian inference different than classical statistics?

* It makes estimates of population parameters from the data.
* It uses methods such as ordinary least squares to determine regression.
* **It maintains probability distributions for population distributions.**
* **Correct**
* **The probability distributions represent quantitative beliefs about their value.**

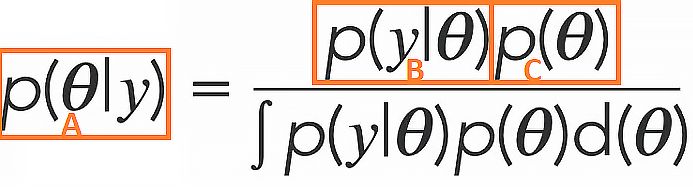
## Question 5 of 6

What statsmodel linear regression property should you use to determine the mean squared error of the residuals?

* mse\_model
* **mse\_resid mse\_resid**
* rsquared

## Question 6 of 6

In Bayes' theorem, as shown here, which highlighted component represents the likelihood?



**B B**

**Correct**

**This is the probability of getting data for given population parameters.**

A