Your grade: 100%

Your latest: **100%** • Your highest: **100%** • To pass you need at least 70%. We keep your highest score.

[Next item →]

1. What type of file saves data in a text-based tabular format? — 1 / 1 point

   - ○ HTML
   - ○ XLSX
   - ⦿ CSV
   - ○ PDF

   ✓ **Correct**
   Correct! A CSV saves data in a text-based tabular format.

2. Which Python libraries are primarily *Algorithmic Libraries*? — 1 / 1 point

   - ○ Jupyter, Regression
   - ⦿ Scikit-learn, Statsmodels
   - ○ Pandas, Numpy, SciPy
   - ○ Matplotlib, Seaborn

   ✓ **Correct**
   Correct! These Python libraries are primarily *Algorithmic Libraries*.

3. What tells you where the data is stored? — 1 / 1 point

   - ⦿ File path
   - ○ Encoding scheme
   - ○ Data format
   - ○ File format

   ✓ **Correct**
   Correct! The file path tells you where the data is stored.

4. What does the `df.tail()` method return? — 1 / 1 point

   - ○ It returns the last five columns of the data frame
   - ○ It returns the first five columns of the data frame
   - ○ It returns the first five rows of the data frame
   - ⦿ It returns the last five rows of the data frame

   ✓ **Correct**
   Correct! The `df.tail()` method returns the last five rows of the data frame.

5. What is a header? — 1 / 1 point

   - ○ The first value in a row
   - ○ The first value in a column
   - ⦿ The name of the columns
   - ○ The name of the rows

   ✓ **Correct**
   Correct! The header refers to the names of the columns.

6. The Scikit-learn library is mostly used for what? — 1 / 1 point

   - ⦿ Machine learning algorithms
   - ○ Data analysis
   - ○ Data visualization

Data visualization

○ Statistical modeling

✓ **Correct**
Correct! The Scikit-learn library is mostly used for machine learning algorithms.

---

7. What is the output of the following code segment of the data frame **df**?

`df.head(5)`

○ It returns all of the rows of the data frame

○ It returns the last five rows of the data frame

○ It returns the header of the data frame

◉ It returns the first five rows of the data frame

1 / 1 point

✓ **Correct**
Correct! The code `df.head(5)` returns the first five rows of data frame **df.**

---

8. What function should you use to remove rows and columns with null or NaN values?

◉ `dropna()`

○ `replacena()`

○ `findna()`

○ `removena()`

1 / 1 point

✓ **Correct**
Correct! The `dropna()` method removes rows and columns with null or NaN values.

---

9. Which type of plot is binning best suited to graph?

○ Scatter plot

○ Line plot

◉ Histogram

○ Box plot

1 / 1 point

✓ **Correct**
Correct! The Histogram plot is best suited for binning to graph.

---

10. What is the primary purpose of standardizing a set of values?

○ To find how well a data set fits a model.

○ To see how many standard deviations each value is from the mean.

○ So you can see the spread of the data set and identify outliers.

◉ It places different variables on the same scale, allowing you to compare them more easily.

1 / 1 point

✓ **Correct**
Correct! Standardizing values serves to place different variables on the same scale, allowing you to compare them more easily.

---

11. Since most statistical models cannot take objects or strings as inputs, what action needs to be performed?

◉ Convert categorical variables into numerical values

○ Convert numerical values into categorical variables

○ Convert object data types into numeric data types

○ Convert numeric data types into object data types

1 / 1 point

✓ **Correct**
Correct! Most statistical models cannot take objects or strings as inputs so it is helpful to convert categorical variables into numerical values.

---

12. What segment of code calculates the mean of the column 'peak-rpm'?

1 / 1 point

○ `df.mean(['peak-rpm'])`

○ `mean = df.['peak-rpm']`

○ `mean( df['peak-rpm'])`

◉ `df['peak-rpm'].mean()`

✓ **Correct**
Correct! This segment of code calculates the mean of the column 'peak-rpm'.

---

13. What does a positive linear relationship between an input variable and an output variable imply?   **1 / 1 point**

○ The output does not adequately explain the input.

○ That as the input increases, the output decreases at about the same rate.

◉ That as the input increases, the output increases at about the same rate.

○ That as the input increases, the output increases at an ever-increasing rate.

✓ **Correct**
Correct! A positive linear relationship between an input variable and an output variable implies that as the input increases, the output increases at about the same rate.

---

14. What is the interquartile range of a data set?   **1 / 1 point**

○ The middle of the data

○ The difference in the range of values in the uppermost quartile with the range of values in the lower-most quartile

◉ The data between the upper and lower quartiles represents the interquartile range.

○ The range of the data, split into four equal-sized groups

✓ **Correct**
Correct! The data between the upper and lower quartiles represents the interquartile range.

---

15. If the predicted function is: **1 / 1 point**

$$\hat{y} = b_0 + b_1 x$$

The method is:

○ Polynomial Regression

○ Exponential Regression

◉ Linear regression

○ Multiple Linear Regression

✓ **Correct**
Correct! A linear regression has 2 parameters and no exponent on the input variable.

---

16. Which of the following statements is true about a model estimator?   **1 / 1 point**

○ Has an order greater than one.

◉ It relates one or more independent variables to an output variable.

○ Has slope and intercept parameters.

○ Has multiple input variables.

✓ **Correct**
Correct! A model estimator relates one or more independent variables to an output variable.

---

17. How are residuals calculated?   **1 / 1 point**

○ $x - \hat{x}$

○ $b_0 + b_1 x$

◉ $y - \hat{y}$

○ $\hat{y} - \bar{y}$

✓ **Correct**
Correct! You calculate residuals by subtracting the predicted value from the actual value.

18. Say you have multiple polynomials that seem to fit your data. What type of graph can you use to help determine which order polynomial is the best fit for your data?

1 / 1 point

○ The order of the polynomials on the x-axis, and the MSE on the y-axis with your training data set.

◉ The order of the polynomials on the x-axis, and the MSE on the y-axis with your testing data set.

○ The input values of your training data on the x-axis and their MSE values on the y-axis.

○ The input values of your testing data on the x-axis and their MSE values on the y-axis.

✓ **Correct**
Correct! A graph of MSE vs. order of the testing data set will help you determine which order polynomial is the best fit for your data.

19. Say you have several differently ordered polynomial models. Which of the following statistics will best help you decide which model to use?

1 / 1 point

◉ Alpha

○ Coefficient of determination

○ Correlation coefficient

○ Mean-squared error

✓ **Correct**
Correct! Ridge regression helps you determine if your model is overfit or underfit.

20. What does the `GridSearchCV()` method do?

1 / 1 point

○ It selects the appropriate hyperparameters for your model.

○ It gives you R2 values for different orders of polynomial models.

◉ It iterates over hyperparameters using cross-validation.

○ It's another way to cross-validate your data set.

✓ **Correct**
Correct! The `GridSearchCV()` iterates over hyperparameters using cross-validation.