

1.

Which scaling approach converts features to standard normal variables?

1 / 1 point

☐

MinMax scaling

☐

Robust scaling

☒

Standard scaling

☐

Nearest neighbor scaling
- ✔

Correct

Correct. Standard scaling converts variables to standard normal variables.

2.

Which variable transformation should you use for ordinal data?

1 / 1 point

☒

Ordinal encoding

☐

One-hot encoding

☐

Min-max scaling☐

✔

Correct

Correct. Use ordinal encoding if there is some order to the categorical features.

3.

What are polynomial features?

1 / 1 point

☐

They are lower order relationships in the data.

☒

They are higher order relationships in the data.

☐

They are logistic regression coefficients.☐

✔

Correct

Correct. Polynomial features are estimated by higher order polynomials in a linear model, like squared, cubed, etc.

4.

What does Boxcox transformation do?

1 / 1 point

☒

It transforms the data distribution into more symmetrical bell curve

☐

It transforms categorical variables into numerical variables.

☐

It makes the data more left skewed☐

✔

Correct

Correct. Boxcox is one of the ways we can transform our skewed dataset to be more normally distributed.

5.

Select three important reasons why EDA is useful.

1 / 1 point

☒

To determine if the data makes sense, to determine whether further data cleaning is needed, and to help identify patterns and trends in the data

☐

To analyze data sets, to determine the main characteristics of data sets, and to use sampling to examine data

☐

To examine correlations, to sample from dataframes, and to train models on random samples of data☐

✔

Correct

Correct. EDA helps us analyze data to summarize its main characteristics.

6.

What assumption does the linear regression model make about data?

1 / 1 point

☐

This model assumes a transformation of each parameter to a linear relationship.

☐

This model assumes an addition of each one of the model parameters multiplied by a coefficient.☐

☒

This model assumes a linear relationship between predictor variables and outcome variables.

✔

Correct

Correct. The linear regression model assumes a linear relationship between predictor and outcome variables.

7.

What is skewed data?

1 / 1 point

☐

Raw data that has undergone log transformation.

☐

Data that has a normal distribution.☐

☒

Data that is distorted away from normal distribution; may be positively or negatively skewed.

✔

Correct

Correct. Often raw data, both the features and the outcome variable, can be negatively or positively skewed.

8.

Select the two primary types of categorical feature encoding.

1 / 1 point

☒

One-hot encoding and ordinal encoding

☐

Encoding and scaling

☐

Frequency encoding and label encoding☐

✔

Correct

Correct. Encoding that transforms non-numeric values to numeric values is often applied to categorical features.

9.

Which scaling approach puts values between zero and one?

1 / 1 point

☐

Nearest neighbor scaling

☐

Standard scaling☐

☒

Min-max scaling

✔

Correct

Correct. Min-max scaling converts variables to continuous variables in the (0, 1) interval by mapping minimum values to 0 and maximum values to 1.

10.

Which variable transformation should you use for nominal data with multiple different values within the feature?

1 / 1 point

☐

Standard scaling

☐

Ordinal encoding☐

☒

One-hot encoding

✔

Correct

Correct. Use one-hot encoding if there are multiple different values within a feature.