Question #1 of 18

An analyst has run several regressions hoping to predict stock returns, and wants to translate this into an economic interpretation for his clients.

$$\text{Return} = 0.03 + 0.020\text{Beta} - 0.0001\text{MarketCap (in billions)} + \varepsilon$$

A correct interpretation of the regression *most likely* includes:

A) a billion dollar increase in market capitalization will drive returns down by 0.01%.

B) prediction errors are always on the positive side.

C) a stock with zero beta and zero market capitalization will return precisely 3.0%.

Question #2 of 18

Which of the following is NOT a model that has a qualitative dependent variable?

A) Logit.

B) Event study.

C) Discriminant analysis.

Question #3 of 18

When utilizing a proxy for one or more independent variables in a multiple regression model, which of the following errors is *most likely* to occur?

A) Model misspecification.

B) Multicollinearity.

C) Heteroskedasticity.

Question #4 of 18

Which of the following is *least likely* to result in misspecification of a regression model?

A) Measuring independent variables with errors.

B) Using a lagged dependent variable as an independent variable.

C) Transforming a variable.

Question #5 of 18

Mary Steen estimated that if she purchased shares of companies who announced restructuring plans at the announcement and held them for five days, she would earn returns in excess of those expected from the market model of 0.9%. These returns are statistically significantly different from zero. The model was estimated without transactions costs, and in reality these would approximate 1% if the strategy were effected. This is an example of:

A) statistical significance, but not economic significance.

B) a market inefficiency.

C) statistical and economic significance.

Question #6 of 18

When constructing a regression model to predict portfolio returns, an analyst runs a regression for the past five year period. After examining the results, she determines that an increase in interest rates two years ago had a significant impact on portfolio results for the time of the increase until the present. By performing a regression over two separate time periods, the analyst would be attempting to prevent which type of misspecification?

A) Incorrectly pooling data.

B) Forecasting the past.

C) Using a lagged dependent variable as an independent variable.

Question #7 of 18

Which of the following is *least likely* to be step in model training?

A) Specify the hyperparameters.

B) Data collection.

C) Evaluation using a performance parameter in the validation sample.

Question #8 of 18

A random forest is *least likely* to:

A) be a classification tree.

B) reduce signal-to-noise ratio.

C) provide a solution to overfitting problem.

Question #9 of 18

What is the main difference between probit models and typical dummy variable models?

A) Dummy variable regressions attempt to create an equation to classify items into one of two categories, while probit models estimate a probability.

B) There is no difference--a probit model is simply a special case of a dummy variable regression.

C) A dummy variable represents a qualitative independent variable, while a probit model is used for estimating the probability of a qualitative dependent variable.

Question #10 of 18

Which of the following statements about supervised learning is *most accurate*?

A) Supervised learning requires human intervention in machine learning process.

B) Typical data analytics tasks for supervised learning include classification and prediction.

C) Supervised learning does not differentiate between tag and features.

Question #11 of 18

Which of the following about unsupervised learning is *most accurate*?

A) Unsupervised learning has lower forecasting accuracy as compared to supervised learning.

B) There is no tag variable.

C) Classification is an example of unsupervised learning algorithm.

Question #12 of 18

What is the appropriate remedy in the presence of excessive number of features in a data set?

A) Dimension reduction.

B) Big data analysis.

C) Unsupervised learning.

Question #13 of 18

Overfitting is *least likely* to result in:

A) higher number of features included in the data set.

B) higher forecasting accuracy in out-of-sample data.

C) inclusion of noise in the model.

Question #14 of 18

A high-yield bond analyst is trying to develop an equation using financial ratios to estimate the probability of a company defaulting on its bonds. Since the analyst is using data over different economic time periods, there is concern about whether the variance is constant over time. A technique that can be used to develop this equation is:

A) logit modeling.

B) multiple linear regression adjusting for heteroskedasticity.

C) dummy variable regression.

Question #15 of 18

Which of the following questions is *least likely* answered by using a qualitative dependent variable?

A) Based on the following executive-specific and company-specific variables, how many shares will be acquired through the exercise of executive stock options?

B) Based on the following company-specific financial ratios, will company ABC enter bankruptcy?

C) Based on the following subsidiary and competition variables, will company XYZ divest itself of a subsidiary?

Question #16 of 18

An analyst is building a regression model which returns a qualitative dependant variable based on a probability distribution. This is *least likely* a:

A) probit model.

B) logit model.

C) discriminant model.

Question #17 of 18

Dimension reduction is an example of:

A) supervised learning.

B) unsupervised learning.

C) clustering.

Question #18 of 18

Which supervised learning model is *most appropriate* (1) when the Y-variable is continuous and (2) when the Y-variable is categorical

	Continuous Y-variable	Categorical Y-variable
A)	Decision trees	Regression
B)	Classification	Neural Networks
C)	Regression	Classification