1. Which of the following is the probability calculus of beliefs, given that beliefs follow certain rules?  
a) Bayesian probability  
b) Frequency probability  
c) Frequency inference  
d) Bayesian inference  
View Answer

Answer: a  
Explanation: Data scientists tend to fall within shades of gray of these and various other schools of inference.

2. Point out the correct statement.  
a) Bayesian inference is the use of Bayesian probability representation of beliefs to perform inference  
b) NULL is the standard missing data marker used in S  
c) Frequency inference is the use of Bayesian probability representation of beliefs to perform inference  
d) None of the mentioned  
View Answer

Answer: a  
Explanation: Frequency probability is the long run proportion of times an event occurs in independent, identically distributed repetitions.

3. Which of the following can be considered as random variable?  
a) The outcome from the roll of a die  
b) The outcome of flip of a coin  
c) The outcome of exam  
d) All of the mentioned  
View Answer

Answer: d  
Explanation: The probability distribution of a discrete random variable is a list of probabilities associated with each of its possible values.

4. Which of the following random variable that take on only a countable number of possibilities?  
a) Discrete  
b) Non Discrete  
c) Continuous  
d) All of the mentioned  
View Answer

Answer: a  
Explanation: Continuous random variable can take any value on some subset of the real line.

5. Point out the wrong statement.  
a) A random variable is a numerical outcome of an experiment  
b) There are three types of random variable  
c) Continuous random variable can take any value on the real line  
d) None of the mentioned  
View Answer

Answer: b  
Explanation: There are two types of random variable-continuous and discrete.

6. Which of the following is also referred to as random variable?  
a) stochast  
b) aleatory  
c) eliette  
d) all of the mentioned  
View Answer

Answer: b  
Explanation: Random variable is also known as stochastic variable.

7. Bayesian inference uses frequency interpretations of probabilities to control error rates.  
a) True  
b) False  
View Answer

Answer: b  
Explanation: Frequency inference uses frequency interpretations of probabilities to control error rates.

8. Which of the following condition should be satisfied by function for pmf?  
a) The sum of all of the possible values is 1  
b) The sum of all of the possible values is 0  
c) The sum of all of the possible values is infinite  
d) All of the mentioned  
View Answer

Answer: a  
Explanation: A probability mass function evaluated at a value corresponds to the probability that a random variable takes that value.

9. Which of the following function is associated with a continuous random variable?  
a) pdf  
b) pmv  
c) pmf  
d) all of the mentioned  
View Answer

Answer: a  
Explanation: pdf stands for probability density function.

10. Statistical inference is the process of drawing formal conclusions from data.  
a) True  
b) False  
View Answer

Answer: a  
Explanation: Statistical inference requires navigating the set of assumptions and tools.

1. Which of the following is the correct formula for total variation?  
a) Total Variation = Residual Variation – Regression Variation  
b) Total Variation = Residual Variation + Regression Variation  
c) Total Variation = Residual Variation \* Regression Variation  
d) All of the mentioned  
View Answer

Answer: b  
Explanation: The complementary part of the total variation is called unexplained or residual.

2. Point out the correct statement.  
a) A standard error is needed to create a prediction interval  
b) The prediction interval must incorporate the variability in the data around the line  
c) Investors use the residual variance to measure the accuracy of their predictions on the value of an asset  
d) All of the mentioned  
View Answer

Answer: d  
Explanation: In statistics, explained variation measures the proportion to which a mathematical model accounts for the variation of a given data set.

3. Which of the following things can be accomplished with linear model?  
a) Flexibly fit complicated functions  
b) Uncover complex multivariate relationships  
c) Build accurate prediction models  
d) All of the mentioned  
View Answer

Answer: d  
Explanation: Linear models are the single most important applied statistical and machine learning technique.

4. Which of the following statement is incorrect with respect to outliers?  
a) Outliers can have varying degrees of influence  
b) Outliers can be the result of spurious or real processes  
c) Outliers cannot conform to the regression relationship  
d) None of the mentioned  
View Answer

Answer: c  
Explanation: Outliers can conform to the regression relationship.

5. Point out the wrong statement.  
a) The fraction of variance unexplained is an established concept in the context of linear regression  
b) “Explained variance” is routinely used in principal component analysis  
c) The general linear model extends simple linear regression (SLR) by adding terms linearly into the model  
d) None of the mentioned  
View Answer

Answer: d  
Explanation: Linearity refers to a mathematical relationship or function that can be graphically represented as a straight line.

6. Which of the following can be useful for diagnosing data entry errors?  
a) hat values  
b) dffit  
c) resid  
d) all of the mentioned  
View Answer

Answer: a  
Explanation: resid returns the ordinary residuals.

7. Multivariate regression estimates are exactly those having removed the linear relationship of the other variables from both the regressor and response.  
a) True  
b) False  
View Answer

Answer: a  
Explanation: Multivariate Data Analysis refers to any statistical technique used to analyze data that arises from more than one variable.

8. Residual \_\_\_\_\_\_ plots investigate normality of the errors.  
a) RR  
b) PP  
c) QQ  
d) None of the mentioned  
View Answer

Answer: c  
Explanation: Patterns in your residual plots generally indicate some poor aspect of model fit.

9. Which of the following show residuals divided by their standard deviations?  
a) rstudent  
b) cooks.distance  
c) rstandard  
d) all of the mentioned  
View Answer

Answer: c  
Explanation: rstandard stands for standardized residuals.

10. The least squares estimate for the coefficient of a multivariate regression model is exactly regression through the origin with the linear relationships.  
a) True  
b) False  
View Answer

Answer: b  
Explanation: Multivariate regression adjusts a coefficient for the linear impact of the other variables.

1. Predicting with trees evaluate \_\_\_\_\_\_\_\_\_\_\_\_\_ within each group of data.  
a) equality  
b) homogeneity  
c) heterogeneity  
d) all of the mentioned  
View Answer

Answer: b  
Explanation: Predicting with trees is easy to interpret.

2. Point out the wrong statement.  
a) Training and testing data must be processed in different way  
b) Test transformation would mostly be imperfect  
c) The first goal is statistical and second is data compression in PCA  
d) All of the mentioned  
View Answer

Answer: a  
Explanation: Training and testing data must be processed in same way.

3. Which of the following method options is provided by train function for bagging?  
a) bagEarth  
b) treebag  
c) bagFDA  
d) all of the mentioned  
View Answer

Answer: d  
Explanation: Bagging can be done using bag function as well.

4. Which of the following is correct with respect to random forest?  
a) Random forest are difficult to interpret but often very accurate  
b) Random forest are easy to interpret but often very accurate  
c) Random forest are difficult to interpret but very less accurate  
d) None of the mentioned  
View Answer

Answer: a  
Explanation: Random forest is top performing algorithm in prediction.

5. Point out the correct statement.  
a) Prediction with regression is easy to implement  
b) Prediction with regression is easy to interpret  
c) Prediction with regression performs well when linear model is correct  
d) All of the mentioned  
View Answer

Answer: d  
Explanation: Prediction with regression gives poor performance in non linear settings.

6. Which of the following library is used for boosting generalized additive models?  
a) gamBoost  
b) gbm  
c) ada  
d) all of the mentioned  
View Answer

Answer: a  
Explanation: Boosting can be used with any subset of classifier.

7. The principal components are equal to left singular values if you first scale the variables.  
a) True  
b) False  
View Answer

Answer: b  
Explanation: The principal components are equal to left singular values if you first scale the variables.

8. Which of the following is statistical boosting based on additive logistic regression?  
a) gamBoost  
b) gbm  
c) ada  
d) mboost  
View Answer

Answer: a  
Explanation: mboost is used for model based boosting.

9. Which of the following is one of the largest boost subclass in boosting?  
a) variance boosting  
b) gradient boosting  
c) mean boosting  
d) all of the mentioned  
View Answer

Answer: b  
Explanation: R has multiple boosting libraries.

10. PCA is most useful for non linear type models.  
a) True  
b) False  
View Answer

Answer: b  
Explanation: PCA is most useful for linear type models.