

## 1. Python\_worksheet\_set\_1

1. Which of the following operators is used to calculate remainder in a division?  
A) #  
B) &  
**C) %**  
D) \$
2. In python 2//3 is equal to?  
A) 0.666  
**B) 0**  
C) 1  
D) 0.67
3. In python, 6<<2 is equal to?  
A) 36  
B) 10  
**C) 24**  
D) 45
4. In python, 6&2 will give which of the following as output?  
**A) 2**  
B) True  
C) False  
D) 0
5. In python, 6|2 will give which of the following as output?  
A) 2  
B) 4  
C) 0  
**D) 6**
6. What does the finally keyword denotes in python?  
A) It is used to mark the end of the code  
**B) It encloses the lines of code which will be executed if any error occurs while executing the lines of code in the try block.**  
C) the finally block will be executed no matter if the try block raises an error or not.  
D) None of the above
7. What does raise keyword is used for in python?  
**A) It is used to raise an exception.**  
B) It is used to define lambda function.  
C) it's not a keyword in python.  
D) None of the above
8. Which of the following is a common use case of yield keyword in python?  
A) in defining an iterator  
B) while defining a lambda function  
**C) in defining a generator**  
D) in for loop

### Multiple correct answer.

9. Which of the following are the valid variable names?  
**A) abc B) 1abc C) abc2** D) None of the above
10. Which of the following are the keywords in python?  
**A) yield B) raise** C) look-in D) all of the above

## 2. Machine Learning Worksheet\_set\_1

1. Which of the following methods do we use to find the best fit line for data in Linear Regression?  
**A) Least Square Error**  
B) Maximum Likelihood  
C) Logarithmic Loss  
D) Both A and B
2. Which of the following statement is true about outliers in linear regression?  
**A) Linear regression is sensitive to outliers**  
B) linear regression is not sensitive to outliers  
C) Can't say  
D) none of these
3. A line falls from left to right if a slope is \_\_\_\_\_?  
A) Positive  
**B) Negative**  
C) Zero  
D) Undefined
4. Which of the following will have symmetric relation between dependent variable and independent variable?  
A) Regression  
**B) Correlation**  
C) Both of them  
D) None of these
5. Which of the following is the reason for over fitting condition?  
A) High bias and high variance  
B) Low bias and low variance  
**C) Low bias and high variance**  
D) None of these
6. If output involves label then that model is called as:  
A) Descriptive model  
B) Predictive modal  
**C) Reinforcement learning**  
D) All of the above
7. Lasso and Ridge regression techniques belong to \_\_\_\_\_?  
A) Cross validation  
B) Removing outliers  
C) SMOTE  
**D) Regularization**
8. To overcome with imbalance dataset which technique can be used?  
**A) Cross validation**  
B) Regularization  
C) Kernel  
D) SMOTE
9. The AUC Receiver Operator Characteristic (AUCROC) curve is an evaluation metric for binary classification problems. It uses \_\_\_\_\_ to make graph?  
**A) TPR and FPR**  
B) Sensitivity and precision  
C) Sensitivity and Specificity  
D) Recall and precision
10. In AUC Receiver Operator Characteristic (AUCROC) curve for the better model area under the curve should be less.  
A) True      **B) False**
11. Pick the feature extraction from below:  
**A) Construction bag of words from a email**  
**B) Apply PCA to project high dimensional data**  
C) Removing stop words

D) Forward selection

### Multiple correct answers.

12. Which of the following is true about Normal Equation used to compute the coefficient of the Linear Regression?

A) We don't have to choose the learning rate.

B) It becomes slow when number of features is very large.

C) We need to iterate.

D) It does not make use of dependent variable.

### Descriptive answers.

13. Explain the term regularization?

Ans- Regularization is a technique used to reduce the errors by fitting the function appropriately on the given training set and avoid overfitting.  
It is a technique that improves model accuracy as well as prevents the loss of important data due to underfitting.

14. Which particular algorithms are used for regularization?

The commonly used regularization techniques are :

1. L1 regularization

2. L2 regularization

3. Dropout regularization

- A regression model which uses **L1 Regularization** technique is called **LASSO(Least Absolute Shrinkage and Selection Operator)** regression.
- A regression model that uses **L2 regularization** technique is called **Ridge regression**. **Lasso Regression** adds "*absolute value of magnitude*" of coefficient as penalty term to the loss function (L).
- **Ridge regression** adds "*squared magnitude*" of coefficient as penalty term to the loss function (L).
- Dropout is a regularization technique **for reducing overfitting in neural networks by preventing complex co-adaptations on training data**.

15. Explain the term error present in linear regression equation?

- Error term is the difference between the expected price at a particular time and the price that was actually observed.
- Error Term Formula:  
$$R^B = \beta_0 + \beta_1 E^x + \epsilon$$
 where  $\beta_0$  and  $\beta_1$  are constants and  $\epsilon$  is an (non constant) error term.
- The error term includes everything that separates your model from actual reality.

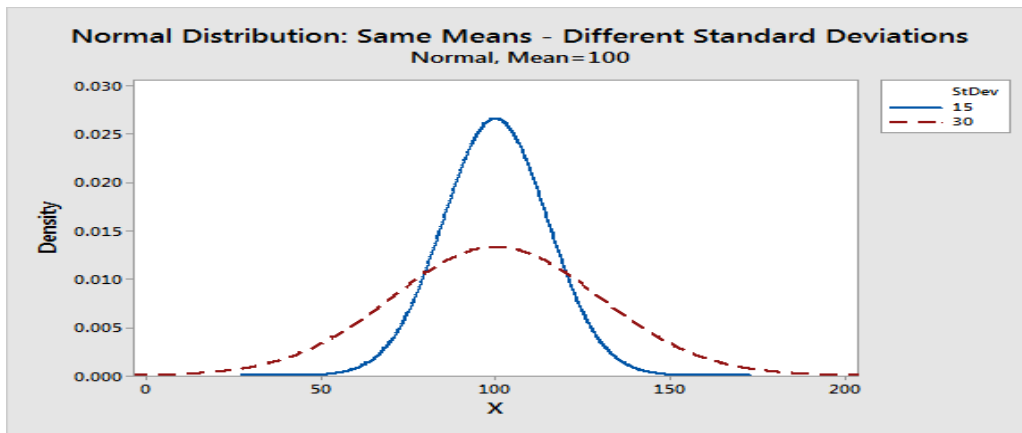
### 3.Statistics Worksheet\_set\_1

1. Bernoulli random variables take (only) the values 1 and 0.  
a) **True**      b) False
2. Which of the following theorem states that the distribution of averages of iid variables, properly normalized, becomes that of a standard normal as the sample size increases?  
a) **Central Limit Theorem**  
b) Central Mean Theorem  
c) Centroid Limit Theorem  
d) All of the mentioned
3. Which of the following is incorrect with respect to use of Poisson distribution?  
a) Modeling event/time data  
b) **Modeling bounded count data**  
c) Modeling contingency tables  
d) All of the mentioned
4. Point out the correct statement.  
a) The exponent of a normally distributed random variables follows what is called the log- normal distribution  
b) Sums of normally distributed random variables are again normally distributed even if the variables are dependent.  
c) The square of a standard normal random variable follows what is called chi-squared distribution  
d) **All of the mentioned**
5. \_\_\_\_\_ random variables are used to model rates.  
a) Empirical  
b) Binomial  
c) **Poisson**  
d) All of the mentioned
6. Usually replacing the standard error by its estimated value does change the CLT.  
a) True      b) **False**
7. Which of the following testing is concerned with making decisions using data?  
a) Probability  
b) **Hypothesis**  
c) Causal  
d) None of the mentioned
8. Normalized data are centered at \_\_\_\_\_ and have units equal to standard deviations of the original data.  
a) **0**  
b) 5  
c) 1  
d) 10
9. 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.

### Descriptive Answers.

10. What do you understand by the term Normal Distribution?

Ans- Normal Distribution is the probability distribution that plots all of its values in a symmetrical fashion, and most of the results are situated around the probability's mean. Values are equally likely to plot either above or below the mean.



11. How do you handle missing data? What imputation techniques do you recommend?

- Deleting Rows with missing values.
- Using Regression Method.
- Impute missing values for continuous variable.
- Impute missing values for categorical variable.
- Using Algorithms that support missing values.
- Prediction of missing values.

12. What is A/B testing?

- It is a way to compare the two versions of a variable to find out which performs better in a controlled environment.
- A/B tests let you evaluate the impact of changes that are relatively inexpensive to implement.
- A/B testing is not only cost effective, it's time efficient.
- A/B testing to make your website as effective as it can be.

13. Is mean imputation of missing data acceptable practice?

- Imputing the mean preserves the mean of the observed data. So if the data are missing completely at random, the estimate of the mean remains unbiased.
- Imputing the mean makes us able to keep our sample size up to the full sample size.

14. What is linear regression in statistics?

- Linear regression is the practice of statistically calculating a straight line that demonstrates a relationship between two different items.
- Linear regression is a linear approach for modelling the relationship between a scalar response and one or more explanatory variables.

15. What are the various branches of statistics?

There are three real branches of statistics:

- **Data collection-** Data collection is a process of gathering information from all the relevant sources to find a solution to the research problem.
- **Descriptive statistics-** Descriptive statistics is the type of analysis of data that helps describe, show or summarize data points in a constructive way such that patterns might emerge that fulfill every condition of the data.
- **Inferential statistics-** Inferential statistics takes data from a sample and makes inferences about the larger population from which the sample was drawn.