**Python Worksheet1**

Q.1 – C

Q.2 - B (Solution is 2/3=0.6666, 2//3=0)

Q.3 - C (shift operators << new bits on the right hand side will be zero, so when 6<<2, 6+6=12\*2=24)

Q.4 – A

Q.5 – B

Q.6 – C

Q.7 – A

Q.8 – C

Q.9 – A & C

Q. 10 – A & B

**Q.11 –**

a = int(input ("Enter a number:"))

f = 1

for i in range(1,a+1):

f = f \* i

print(f'The factorial of the number is{f}')

**Q.12**

import math

n = int(input())

if n < 2:

print("A number must be 2 or more")

quit()

elif n == 2:

print("It is a prime number")

quit()

else:

print("It is not a prime number")

**Q. 13**

string= input(("Enter a string"))

if string == string[::-1]:

print("The string is a palindrome")

else:

print("Not a palindrome")

**Q.14**

import math

a=int(input("Enter first side: "))

b=int(input("Enter second side: "))

c=int(input("Enter third side: "))

s=(a+b+c)/2

area=math.sqrt(s\*(s-a)\*(s-b)\*(s-c))

print("Area of the triangle is: ",round(area,2))

**Q.15**

test\_str = string

all\_str = {}

for i in test\_str:

if i in all\_str:

all\_str[i]+=1

else:

all\_str[i]=1

print('Count of all characters in string is\n'+str(all\_str))

**Statistics Worksheet 1**

Q. 1 – A

Q. 2 – A

Q. 3 – B

Q. 4 – D

Q. 5 – C

Q.6 – B

Q.7 – C

Q.8 – A

Q.9 – C

Q.10 – Normal distribution, also known as the Gaussian distribution, is a probability distribution that is symmetric about the mean, showing that data near the mean are more frequent in occurrence than data far from the mean. In graph form, normal distribution will appear as a bell curve.

Q.11 – Data may be missed due to **Missing Completely At Random (MCAR), Missing At Random (MAR) and Not Missing At Random (NMAR) however we can follow these imputation Mean or Median Imputation, Multivariate Imputation by Chained Equations (MICE) and Random Forest**

Q.12 – A/B testing is a basic randomized control experiment. It is a way to compare the two versions of a variable to find out which performs better in a controlled environment. For instance, let's say you own a company and want to increase the sales of your product.

Q.13 – Bad practice in general If just estimating means: mean imputation preserves the mean of the observed data. Leads to an underestimate of the standard deviation. Distorts relationships between variables by “pulling” estimates of the correlation toward zero

Q.14 – Linear regression attempts to model the relationship between two variables by fitting a linear equation to observed data. One variable is considered to be an explanatory variable, and the other is considered to be a dependent variable.

Q.15 – descriptive **statistics** and inferential **statistics,** both of these are employed in scientific analysis of data and both are equally important for the student of **statistics**.

**Machine Learning**

Q.1 – D

Q.2 – C

Q.3 – B

Q.4 – C

Q.5 – A

Q.6 – A

Q.7 – C

Q.8 – C

Q.9 – A,B & C

Q.10 – A & D

Q.11 – C

**Q.12 -**

You could use batch gradient descent, stochastic gradient descent, or mini-batch gradient descent. SGD and MBGD would work the best because neither of them need to load the entire dataset into memory in order to take 1 step of gradient descent. Batch would be ok with the caveat that you have enough memory to load all the data. The normal equations method would not be a good choice because it is computationally inefficient. The main cause of the computational complexity comes from inverse operation on an (n x n) matrix.

**Q.13 -**

The normal equations method does not require normalizing the features, so it remains unaffected by features in the training set having very different scales. Feature scaling is required for the various gradient descent algorithms. Feature scaling will help gradient descent converge quicker.