

Machine learning:

1. A
2. A
3. B
4. B
5. C
6. B
7. D
8. D
9. A
10. B
11. B
12. A & b
13. **Regularization** is a technique that modifies or constrains the complexity of a statistical model to prevent overfitting or underfitting. Overfitting occurs when the model fits the training data too well, but fails to generalize to new or unseen data. Underfitting occurs when the model is too simple or biased, and cannot capture the underlying patterns or relationships in the data.
- 14.
15. **An error term** means that the model is not completely accurate and results in differing results. It accounts for the discrepancies between the observed values of the dependent variable and the values predicted by the regression equation.

Python:

1. C
2. B
3. C
4. A
5. D
6. C
7. A
8. C
9. A & c
10. A & b
11.

```
num = int(input("Enter a number:"))
factorial = 1

if num < 0:
    print ("Factorial does not exist")
elif num == 0:
    print('The factorial is 1')
else:
    for i in range (1,num+1):
        factorial = factorial*i
    print("The factorial of",num,"is",factorial)
```

output...

```
Enter a number:5
The factorial of 5 is 1
The factorial of 5 is 2
The factorial of 5 is 6
The factorial of 5 is 24
The factorial of 5 is 120
```

```
12..num=int(input("Enter the number:"))
count=0
i=1
while i<=num:
    if num%i==0:
        count=count+1
        i=i+1

    if count==2:
        print("Its a Prime Number:")
    elif count>2:
        print("Its a Composite number:")
    else:
        print("The number is neither Prime nor Composite")
```

Output

```
Enter the number:4
The number is neither Prime nor Composite
Its a Prime Number:
```

```
13..n = input("Enter a string:")
if n == n[::-1]:
    print(n+" is a palindrome")
else:
    print(n+ "is not a palindrome")
```

Output

```
Enter a string:naman
naman is a palindrome
```

```
14..import math
```

```
a = float(input("Enter first shorter side of right angled triangle:"))  
b = float(input("enter second shorter side of right angled triangle:"))  
c = math.sqrt(a ** 2+b ** 2)  
print("Hypotenuse of a right angled triangle is : %.2f"%c)
```

Output

```
Enter first shorter side of right angled triangle:10  
enter second shorter side of right angled triangle:5  
Hypotenuse of a right angled triangle is : 11.18
```

```
15..my_string =input("Enter a string:")  
count = {}  
  
for letter in my_string:  
    if letter in count:  
        count[letter] +=1  
    else:  
        count[letter] =1  
  
print("Count frequency is ...")  
for key, value in count.items():  
    print(f"{key} occurs {value} times")
```

Output

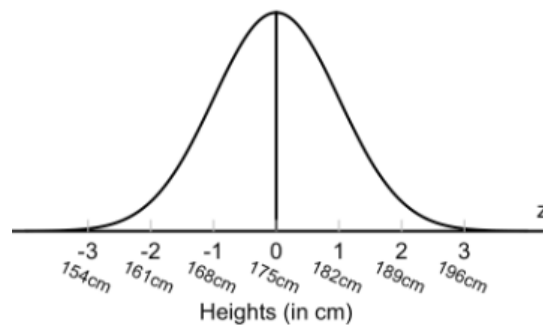
```
Enter a string:count frequency  
Count frequency is ...  
c occurs 2 times  
o occurs 1 times  
u occurs 2 times  
n occurs 2 times  
t occurs 1 times  
   occurs 1 times  
f occurs 1 times  
r occurs 1 times  
e occurs 2 times  
q occurs 1 times  
y occurs 1 times
```

Statistics:

1. A
2. A
3. B
4. D
5. C
6. B
7. B
8. A
9. C

10. What is Normal Distribution?

- It is a type of continuous probability distribution which is symmetrical around its mean
- Majority of the data points bend towards the middle of a range, also called the mean
- It is also known as a *Gaussian distribution* or *probability bell curve*.



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

- In statistics, imputation is the process of replacing missing data with substituted values.
- After this we can analyse the entire data set as if the imputed values were the true observed values
- Data imputation is the process of replacing missing or unavailable entries in a dataset with substituted values. This process is important for maintaining the integrity of data analysis.
- Types
 1. Mean/Median/Mode Imputation: In this process we replace missing values with the mean, median, or mode of the available data points within the dataset.
 2. Random Imputation: This process involves filling in missing values with random observations from the dataset
 3. Hot-deck imputation: This process replaces missing values with observed responses from similar or closest values. A value picked at random from a sample member who has comparable values on other variables.
 4. Regression imputation: This process involves using a regression model to predict missing values based on other available variables.
 5. Cold deck imputation: In this process a value picked deliberately from an individual with similar values on other variables. In most aspects, this is comparable to Hot Deck, but without the random variance

12. What is A/B testing?

- It is also called split testing or bucket testing
- This process involves comparing performance of two versions of content to see which one is preferred more by viewers
- It helps to identify how to provide the best user experience
- Examples
 - Website A/B testing
 - We can use tools such as images, designs, calls to action etc
 - Now the traffic will split between two versions—A and B
 - After this analysis can be done to identify which version yields the highest number of conversions
- A/B testing can be used for analysing any digital marketing tools such as emails, newsletters, messages, web pages etc. It helps in measurement of number of visitors, click-through rates, new signups and subscriptions

13. Is mean imputation of missing data acceptable practice?

- Mean imputation means the process of replacing null values in a data collection with the data's mean.
- It is not considered a desirable practice because
 1. It ignores feature correlation
 2. It decreases the variance of our data while increasing bias, hence the data model becomes less accurate

14. What is Linear Regression in statistics?

- Linear regression analysis is used to predict the value of a variable based on the value of another variable.
- The variable you want to predict is called **the dependent variable**.
- The variable you are using to predict the other variable's value is called the **independent variable**

15. What are the various branches of Statistics?

- There are two main branches of statistics, descriptive statistics and inferential statistics.
- **Descriptive Statistics:**

- It deals with the presentation and collection of data
- The main features of a dataset are summarized and focus is on visualization
- Common methods and techniques are
 - mean, median, and mode
 - range, variance, and standard deviation
 - frequency distributions (histograms, frequency tables)
 - graphical representations (box plots, bar charts, pie charts, etc.)
- **Inferential Statistics**
 - It involves drawing conclusions from the statistical analysis that has been performed This process involves making inferences, predictions, or generalizations about a larger population based on data collected from a sample of that population.