(**ANSWERSHEET 3(MACHINE LEARNING)**

Answers 1 to Answers 8

Ans 1 (C) y- intercept

Ans 2 (A) True

Ans 3 (B) The dependent variable

Ans 4 (B) Linear Regression

Ans 5 (C) The correlation coefficient squared

Ans 6 (B) y increases as x increases

Ans 7 **(C**) both linear and non-linear data

Ans 8 (A) 0 to 1

ANSWERS 9 TO ANSWERS 13

Ans 9 (B) RMSE

(D) MAE

Ans 10 (A) Linear regression is a supervised learning algorithm

(C) Shape of linear regression's cost function is convex.

(D) Linear regression is used to predict discrete dependent variable.

Ans 11 (A) Ridge

(B) Lasso

(D) Elastic Net

Ans 12

Ans 13 (A) Linearity

(B) Homoscedasticity

(D) Normality

<u>ANSWERS 14 & ANSWERS 15</u>

Ans 14

Linear regression is a data analysis technique that predicts the value of unknown data by using another related and known data value. It mathematically models the unknown or dependent variable and the known or independent variable as a linear equation.

Suppose, that we have data about our expenses and income for last year. Linear regression techniques analyze this data and determine that our expenses are half our income. They then calculate an unknown future expense by halving a future known income.

Ans 15

<u>Simple Linear Regression</u>:

The association between two variables is established using a straight line in Simple Linear Regression. It tries to create a line that is as near to the data as possible by determining the slope and intercept, which define the line and reduce regression errors.

There is a single x and y variable

- Equation: Y = mX+c

<u>Multiple Linear Regression</u>:

Multiple linear regressions are based on the presumption that both the dependent and independent variables, or Predictor and Target variables, have a linear relationship.

There are two types of multilinear regressions: linear and nonlinear.

It has one or more x variables and one or more y variables, or one dependent variable and two or more independent variables

Equation: $Y = m1X1+m2X2+m3X3+..c$

Where,

Y = Dependent Variable
m = Slope
X = Independent Variable
c = Intercept.

ANSWERSHEET –3 (PYTHON)

ANSWERS 1 TO ANSWERS 8

Ans 1 (D) int('32')

Ans 2 (C) 4

Ans 3 (B) (a**b)%c

Ans 4 (D) <class 'type'>

Ans 5 (A) 'A'

Ans 6 (D) Method

Ans 7 (B) False

Ans 8 (B) sometimes

ANSWERS 9 AND ANSWER 10

Ans 9 (A) -68.7e100

(B) 42e3

(D) 3.0

Ans 10 B) You can pass keyword arguments in any order.

C) You can call a function with positional and keyword arguments.

D) Positional arguments must be before keyword arguments in a function call.

Ans 11

In [1]:
```
# Python 3.x code to demonstrate star pattern

# Function to demonstrate printing pattern
def pypart(n):

    # outer loop to handle number of rows
    # n in this case
    for i in range(0, n):

        # inner loop to handle number of columns
        # values changing acc. to outer loop
        for j in range(0, i+1):

            # printing stars
            print("* ",end="")

        # ending line after each row
        print("\r")

 # Driver Code
n = 5
pypart(n)
```

```
*
* *
* * *
* * * *
* * * * *
```

Ans 12

In [3]:
```
# Hourglass pattern in Python

# Reading number of rows
row = int(input("Enter number of rows: "))

print("Generated Hourglass Pattern is: ")
# Upper-half
for i in range(row, 0, -1):
    for j in range(row-i):
        print(" ", end="")
    for j in range(1, 2*i):
        print("*", end="")
    print()

# Lower-half
for i in range(2, row+1):
    for j in range(row-i):
        print(" ", end="")
    for j in range(1, 2*i):
        print("*", end="")
    print()
```

```
Enter number of rows: 5
Generated Hourglass Pattern is: 5
*********
 *******
  *****
   ***
    *
   ***
  *****
 *******
*********
```

Ans 13

In [4]:
```
# Print Pascal's Triangle in Python
from math import factorial

# input n
n = 5
for i in range(n):
    for j in range(n-i+1):

        # for left spacing
        print(end=" ")

    for j in range(i+1):

        # nCr = n!/((n-r)!*r!)
        print(factorial(i)//(factorial(j)*factorial(i-j)), end=" ")

    # for new line
    print()
```

```
      1
     1 1
    1 2 1
   1 3 3 1
  1 4 6 4 1
```

Ans 14

In [5]:
```
# Python program to
# print Diamond shape

# Function to print
# Diamond shape
def Diamond(rows):
    n = 0
    for i in range(1, rows + 1):
        # loop to print spaces
        for j in range (1, (rows - i) + 1):
            print(end = " ")

        # loop to print star
        while n != (2 * i - 1):
            print("*", end = "")
            n = n + 1
        n = 0

        # line break
        print()

    k = 1
    n = 1
    for i in range(1, rows):
        # loop to print spaces
        for j in range (1, k + 1):
            print(end = " ")
        k = k + 1

        # loop to print star
        while n <= (2 * (rows - i) - 1):
            print("*", end = "")
            n = n + 1
        n = 1
        print()
  # Driver Code
# number of rows input
rows = 5
Diamond(rows)
```

```
    *
   ***
  *****
 *******
*********
 *******
  *****
   ***
    *
```

In [ ]:

In [ ]:

ANSWERSHEET 3 (STATISTICS)

Answers 1 to Answers 12

Ans 1 (C) Neither

Ans 2 (B) The underlying distribution.

Ans 3 (A) True

Ans 4 (B) We are 95% confident that the results have not occurred by chance.

Ans 5 (C) If the region of rejection is located in one or two tails of the distribution.

Ans 6 (C) We accept a null hypothesis when it is not true.

Ans 7 (A) It is a sample proportion

Ans 8 (A) .013

Ans 9 (C) 1.667

Ans 10 (B) –2.50

Ans 11 (C). There is a difference between the proportions of American men and American women who belong to sports clubs.

Ans 12(B) It is reasonable to say that more than 40% of Americans exercise regularly.

ANSWERS 13 TO ANSWERS 15

Ans 13

A two-sample independent *t*-test can be run on sample data from a normally two distributed outcome variable to determine if its mean differs across two independent groups.

For example, we could see if the mean GPA differs between freshman and senior college students by collecting a sample of each group of students and recording their GPAs.

Hypotheses:

H_o: The population mean of one group equals the population mean of the other group, or $\mu_1 = \mu_2$

$_{HA}$: The population mean of one group does not equal the population mean of the other group, or $\mu_1 \neq \mu_2$

This test can also be conducted with a directional alternate hypothesis:

H_o: The population mean of one group equals the population mean of the other group, or $\mu_1 = \mu_2$

$_{Ha}$: The population mean of one group is greater than the population mean of the other group, or $\mu_1 > \mu_2$

The test statistic for a two-sample independent *t*-test is calculated by taking the difference in the two sample means and dividing by either the pooled or unpooled estimated standard error. The estimated standard error is an aggregate measure of the amount of variation in both groups.

Degree of freedom : Varies by conditions, but the basic rule of thumb for hand calculations is the smaller of $n_1 - 1$ and $n_2 - 1$, where n is the sample size for each group.

$$t = \frac{\bar{X}_1 - \bar{X}_2}{SE}$$

Assumptions:

- Random samples
- Independent observations
- The population of each group is normally distributed.
- The population variances are equal.

Ans 14

Statistics problems often involve comparisons between two independent sample means.

Suppose we have two populations with means equal to μ_1 and μ_2. Suppose further that we take all possible samples of size n_1 and n_2.

The samples are independent that is, observations in population 1 are not affected by observations in population 2, and vice versa.

The size of each population is large relative to the sample drawn from the population. That is, N_1 is large relative to n_1, and N_2 is large relative to n_2. (In this context, populations are considered to be large if they are at least 20 times bigger than their sample.

The set of differences between sample means is normally distributed. This will be true if each population is normal or if the sample sizes are large. (Based on the central limit theorem, sample sizes of 40 would probably be large enough).

Some assumptions are:

The expected value of the difference between all possible sample means is equal to the difference between population means. Thus,

- $E(x_1 - x_2) = \mu_d = \mu_1 - \mu_2$.

The standard deviation of the difference between sample means (σ_d) is approximately equal to:

- $\sigma_d = \text{sqrt}(\sigma_1^2 / n_1 + \sigma_2^2 / n_2)$

It is straightforward to derive the last bullet point, based on material covered in previous lessons. The derivation starts with a recognition that the variance of the difference between independent random variables is equal to the sum of the individual variances. Thus,

$$\sigma^2_d = \sigma^2_{(x1 - x2)} = \sigma^2_{x1} + \sigma^2_{x2}$$

If the populations N_1 and N_2 are both large relative to n_1 and n_2, respectively, then

$$\sigma^2_{x1} = \sigma^2_1 / n_1$$

$$\sigma^2_{x2} = \sigma^2_2 / n_2$$

$$\sigma_d^2 = \sigma_1^2 / n_1 + \sigma_2^2 / n_2$$

$$\sigma_d = \text{sqrt}(\sigma_1^2 / n_1 + \sigma_2^2 / n_2)$$

Ans 15

The two-sample t-test (Snedecor and Cochran, 1989) is used to determine if two population means are equal. A common application is to test if a new process or treatment is superior to a current process or treatment. There are several variations on this test. The data may either be paired or not paired.

Two sample t-test are

I) Independent two-sample t-test.

II)Paired sample t-test.

Example of two-sample t-test is Comparing the average test scores of two classes from two different schools. Comparing the average weights of two different groups of people. Measuring the difference in height between men and women.