

WORKSHEET 3 - PYTHON

Ans.1)- B) `int(3.2)`

Ans.2)- C) 4

Ans.3)- B) `(a**b)%c`

Ans.4)- A) `< class 'type'>`

Ans.5)- C) 65

Ans.6)- D) Method

Ans.7)- B) False

Ans.8)- B) Sometimes

Ans.9)- A,B,C,D

Ans.10)- B,C,D

Ans.11)- 15)- Link-

<https://github.com/sumeshyadav29/fliprobo/blob/main/Worksheet/W3%20python.ipynb>

MACHINE LEARNING

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)- A) linear data

Ans.8)- A) 0 to 1

Ans.9)- A,D

Ans.10)-A, C

Ans.11)-A,B,D

Ans.12)-A

Ans.13)-A,B,C,D

Ans.14)- Linear regression is a statistical method used to model the relationship between a dependent variable and one or more independent variables. It assumes that the relationship between the dependent variable and the independent variable(s) is linear. The goal of linear regression is to find the line (or hyperplane) that best fits the data by estimating the coefficients that minimize the sum of the squared distances between the observed values and the predicted values. There are two types of linear regression: simple linear

regression and multiple linear regression. Linear regression has many applications, such as predicting sales, estimating housing prices, and predicting student test scores. It is a simple and interpretable method that is widely used in many fields.

Ans.15)- Simple linear regression is a statistical method that models the relationship between a dependent variable and a single independent variable. It assumes that the relationship between the dependent variable and the independent variable is linear. The goal of simple linear regression is to estimate the slope and intercept of the line that best fits the data, which can be used to predict the value of the dependent variable given a new value of the independent variable.

Multiple linear regression is a statistical method that models the relationship between a dependent variable and two or more independent variables. It assumes that the relationship between the dependent variable and the independent variables is linear. The goal of multiple linear regression is to estimate the coefficients of the line (or hyperplane) that best fits the data, which can be used to predict the value of the dependent variable given new values of the independent variables.

STATISTICS

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)- C) -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.

Ans.13)- To find the test statistic for two samples, you can use a t-test or a z-test, depending on whether the population standard deviation is known or unknown. Here are the basic steps for each:

1. For a t-test with unknown population standard deviation, calculate the sample means (\bar{x}_1 and \bar{x}_2), the sample standard

deviations (s_1 and s_2), the sample sizes (n_1 and n_2), and the degrees of freedom ($df = n_1 + n_2 - 2$).

2. Calculate the pooled standard deviation using the formula: $sp = \sqrt{((n_1-1)s_1^2 + (n_2-1)s_2^2) / df}$.
3. Calculate the standard error of the difference between the sample means using the formula: $SE = \sqrt{sp^2/n_1 + sp^2/n_2}$.
4. Calculate the t-statistic using the formula: $t = (x_1 - x_2) / SE$.
5. Look up the t-distribution table or use a statistical software to find the p-value associated with the calculated t-statistic and the degrees of freedom.
6. Compare the p-value to the chosen level of significance (α) to make a decision about the null hypothesis.

For a z-test with known population standard deviation, the steps are similar except that you use the population standard deviation instead of the pooled standard deviation to calculate the standard error of the difference between the sample means. The t-distribution table is replaced with the standard normal distribution table, and the test statistic is calculated using the formula: $z = (x_1 - x_2) / (sd / \sqrt{n})$, where sd is the population standard deviation and n is the sample size.

Ans.14)- To find the sample mean difference, you simply subtract the mean of one sample from the mean of the other sample.

Mathematically, this can be expressed as:

$$\text{sample mean difference} = x_1 - x_2$$

where x_1 and x_2 are the means of the two samples being compared.

Ans.15)- A two-sample t-test is used to compare the means of two independent samples and determine if they are significantly different from each other.