# WORKSHEET 3 PYTHON

1. A
2. C) 4
3. B) (a**b)%c
4. A) <class ‘type’>
5. A) ‘A’
6. D) Method
7. B) False
8. B) Sometimes
9. A), B),C),D)
10. A) You can pass positional arguments in any order

**STATISTICS**

1. c. Neither
2. b. The underlying distribution
3. a. True
4. b. We are 95% confident that the results have not occurred by chance
5. c. If the region of rejection is located in one or two tails of the distribution
6. c. We accept a null hypothesis when it is not true
7. a. It is a sample proportion
8. a. 0.013
9. c. 1.667
10. c. -2.50
11. c. There is a difference between the proportions of American men and American women who belong to sports clubs.
12. b. It is reasonable to say that more than 40% of Americans exercise regularly.
13. Test statistic formula:

$$t=(x'_1-x'_2)/\text{sqrt}[(s_1^2/n_1)+(s_2^2/n_2)]$$

$x'_1$ =mean of $n_1$ independent observations from a normally distributed population with variance $s_1^2$.

$x'_2$= mean of $n_2$ independent observations from a normally distributed population with variance $s_2^2$.

14. Sample mean difference:

$$x'_1-x'_2 =\mu_1-\mu_2$$

$x'_1$ =mean of $n_1$ independent observations from a normally distributed population with mean $\mu_1$ and variance $\sigma_1^2$.

$x'_2$= mean of $n_2$ independent observations from a normally distributed population with mean $\mu_2$ and variance $\sigma_2^2$.

$x'_1 \sim N(\mu_1, \sigma_1^2/n_1)$

$x'_2 \sim N(\mu_2, \sigma_2^2/n_2)$

$x'_1 - x'_2 \sim N(\mu_1-\mu_2, (\sigma_1^2/n_1)-(\sigma_2^2/n_2))$

15. Two Sample t-test is a method used to test whether the means of two population are equal or not.

Example:

Group 1= [15,14,16,12,14,12,10]

Group 2= [18,19,17,16,14,14,18]

Use t-test formula:

$$t=(x'_1-x'_2)/\text{sqrt}[(s_1^2/n_1)+(s_2^2/n_2)]$$

$x'_1=93/7=13.3$

$x'_2=116/7=16.6$

$s_1^2=4.2381$ $s_1=2.0587$

$s_2^2=3.9533$ $s_2=1.9881$

$n_1=7$ $n_2=7$

$t=|-3.0508|$

The probability of t=3.0508 with degree of freedom (df) = (7-1)=6 falls between the one sided p values of .01 and .02 close to .01

p value is less than critical p-value of alpha(.05).so, we reject the null hypothesis.

# MACHINE LEARNING

1. C)
2. A)
3. B)
4. B)
5. C)
6. B)
7. A)
8. A)
9. B) & D)
10. A) & D)
11. A) & B)
12. A) &C)
13. A),B) & D)
14. Linear Regression is the supervised machine learning model in which model finds the linear relationship between the dependent (target) and independent (predictors) variable. It is to find the best fit linear line and optimal values of intercept and coefficients such that the error is minimized.

    There are two types of linear regression-Simple and Multiple.

Simple Linear Regression is only one independent variable and the model has to find the linear relationship of it with the dependent variable.
Multiple Linear Regression is more than one independent variables for the model to find the relationship.
Assumptions of Linear Regression:
1. Linearity: It states that the dependent variable Y should be linearly related to independent variables.
2. Normality: The X and Y variables should be normally distributed
3. Homoscedasticity: The variance of the error terms should be constant i.e., the spread of residuals should be constant for all values of X
4. Independence: The variables should be independent of each other i.e., no correlation should be there between the independent variables

15.

| Simple Linear Regression | Multiple Linear Regression |
|---|---|
| One independent variable | More than one independent variable |
| Always linear relationship with independent and dependent variable | Can be linear and non-linear relationship with independent and dependent variable |
| Y= b0 +b1*x | Y=b0 +b1*x +b2*x +……bn*xn |