**Assignment 16**

1.In a linear equation, what is the difference between a dependent variable and an independent variable?

**Ans:** **The independent variable is the cause.** **Its value is independent of other variables in your study.** **The dependent variable is the effect.** **Its value depends on changes in the independent variable**.

**A linear equation in two variables can be described as a linear relationship between x and y, that is, two variables in which the value of one of them (usually y) depends on the value of the other one (usually x). In this case, x is the independent variable, and y depends on it, so y is called the dependent variable.**

2.What is the concept of simple linear regression? Give a specific example.

**Ans: y = mx+c could be simple linear regression. For ex - We could use the equation to predict weight if we knew an individual's height. In this example, if an individual was 70 inches tall, we would predict his weight to be: Weight = 80 + 2 x (70) = 220 lbs. In this simple linear regression, we are examining the impact of one independent variable on the outcome.**

3. In a linear regression, define the slope.

**Ans: The slope of a regression line (b) represents the rate of change in y as x changes. Because y is dependent on x, the slope describes the predicted values of y given x.**

4. Determine the graph's slope, where the lower point on the line is represented as (3, 2) and the higher point is represented as (2, 2).

**Ans: slope is 0**

5. In linear regression, what are the conditions for a positive slope?

**Ans: In simpler words, a positive slope is one in which the variable x increases with the increase in variable y and/or variable y increases with the increase in variable x. Similarly, the variable x decreases with the decrease in variable y, and/or variable y decreases with the decrease in variable x.**

6. In linear regression, what are the conditions for a negative slope?

**Ans: In other words, a negative slope is one in which the variable x increases with the decrease in variable y and/or variable y increases with the decrease in variable x. In the same manner, the variable x decreases with the increase in variable y, and/or variable y decreases with the increase in variable x.**

7. What is multiple linear regression and how does it work?

**Ans: Multiple linear regression refers to a statistical technique that uses two or more independent variables to predict the outcome of a dependent variable. The technique enables analysts to determine the variation of the model and the relative contribution of each independent variable in the total variance.**

8. In multiple linear regression, define the number of squares due to error.

**Ans : The main reason is that squared error allows to decompose each observed value into the sum of orthogonal components such that the sum of observed squared values is equal to the sum of squared components**.

9. In multiple linear regression, define the number of squares due to regression.

**Ans: The mean square due to regression, denoted MSR, is computed by dividing SSR by a number referred to as its degrees of freedom; in a similar manner, the mean square due to error, MSE, is computed by dividing SSE by its degrees of freedom**.

10. In a regression equation, what is multicollinearity?

**Ans: Multicollinearity occurs when two or more independent variables are highly correlated with one another in a regression model. This means that an independent variable can be predicted from another independent variable in a regression model**.

11. What is heteroskedasticity, and what does it mean?

**Ans: Heteroskedasticity refers to situations where the variance of the residuals is unequal over a range of measured values. When running a regression analysis, heteroskedasticity results in an unequal scatter of the residuals (also known as the error term).**

12. Describe the concept of ridge regression.

**Ans: Ridge regression is a way to create a parsimonious model when the number of predictor variables in a set exceeds the number of observations, or when a data set has multicollinearity (correlations between predictor variables)**

13. Describe the concept of lasso regression.

**Ans: Lasso regression is a regularization technique. It is used over regression methods for a more accurate prediction. This model uses shrinkage. Shrinkage is where data values are shrunk towards a central point as the mean. The lasso procedure encourages simple, sparse models (i.e. models with fewer parameters).**

14. What is polynomial regression and how does it work?

Ans: Polynomial Regression is **a form of Linear regression known as a special case of Multiple linear regression which estimates the relationship as an nth degree polynomial**. Polynomial Regression is sensitive to outliers so the presence of one or two outliers can also badly affect the performance

15. Describe the basis function.

**Ans: A radial basis function (RBF) is a function that assigns a real value to each input from its domain (it is a real-value function), and the value produced by the RBF is always an absolute value; i.e. it is a measure of distance and cannot be negative.**

16. Describe how logistic regression works.

**Ans: Logistic Regression is a “Supervised machine learning” algorithm that can be used to model the probability of a certain class or event. It is used when the data is linearly separable and the outcome is binary or dichotomous in nature. That means Logistic regression is usually used for Binary classification problems.**