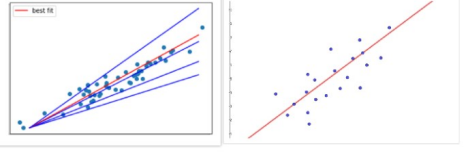
1. What is linear regression?

1. It tries to find out the best possible linear relationship between the input features and the target variable(y).

2. “It is a supervised machine learning algorithm that best fits the data which has the target variable (dependent variable) as a linear combination of the input features (independent variables).”

3. The target variable is also known as an independent variable or label.

4. Input features are also known as dependent variables.



5. When you think of linear regression think of fitting a line such that the distance between the data points and the line is minimum. As shown above, the red line best fits that data than the other blue lines.

6. Linear equation is, y = mx + c

(The goal of the linear regression is to find the best values for m and c that represents the given data.)

1. **How do you represent a simple linear regression?**
2. 'Simple linear regression has one independant variable and one dependant variable,
3. simple linear regression Equation is Y = mx + c

X – will be one independent Variable

Y – will be dependent variable

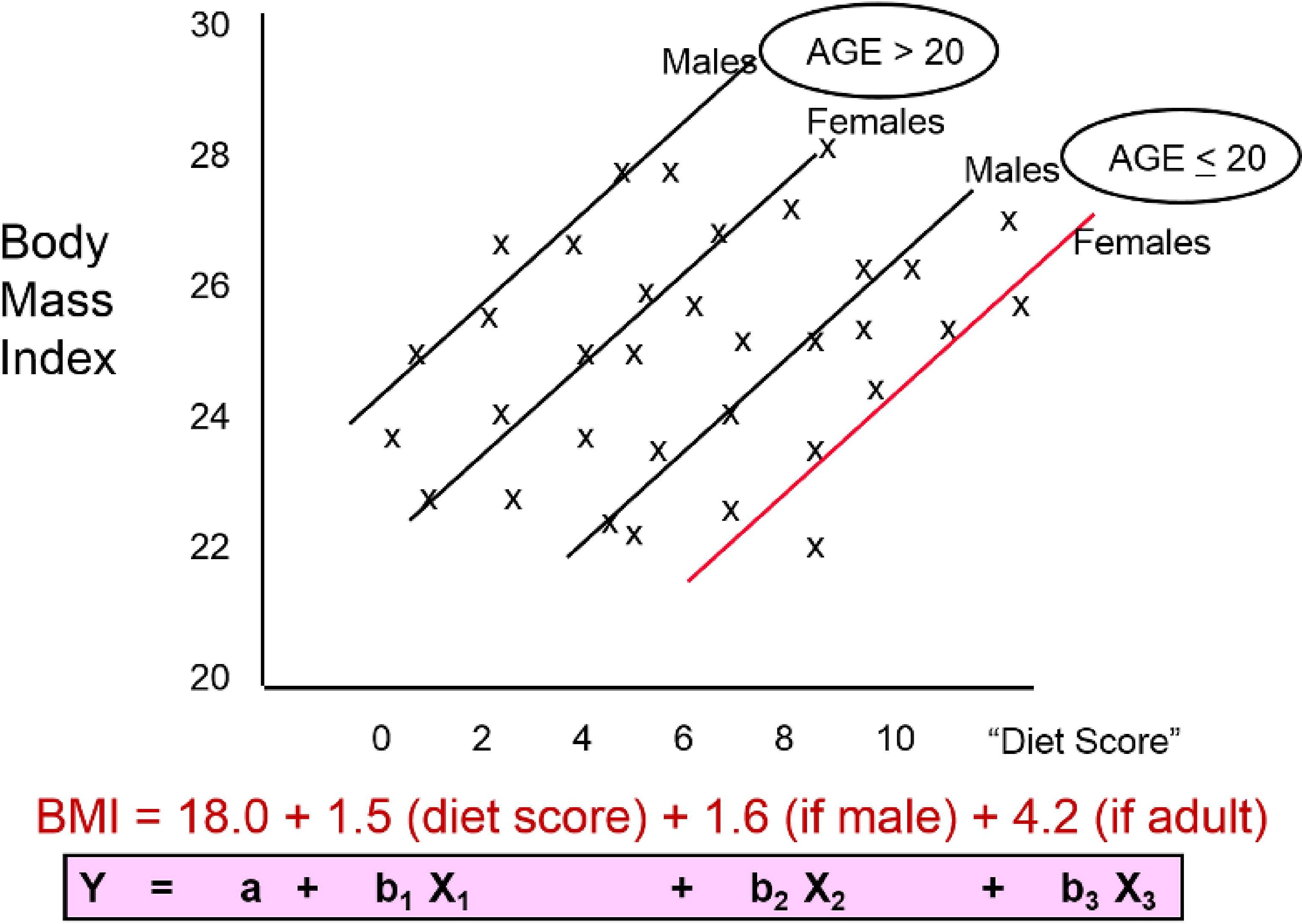
m - is the regression coefficient or slope of line – how much we expect y to change as x increases.

c - is the intercept, the predicted value of y when the x is 0.

1. (e.g. the amount of soil erosion at a certain level of rainfall).
2. What is multiple linear regression?
3. Multiple linear regression is used to estimate the relationship between two or more independent variables and one dependent variable. You can use multiple linear regression when you want to know:
4. We can represent multiple linear regression Equation is y=m1x1 + m2x3 + .... + mnxn + C

M1 m2 m3 are slope of feature or independent variable x1 x2 and x3

Y will be dependent variable

1. 

**4.What are the assumptions made in the Linear regression model?**

1. **Linearity:**

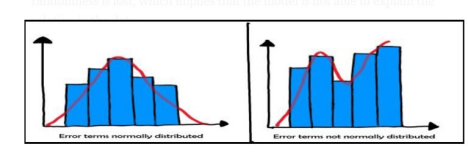
There should be linear relationship between dependent and independent variables.

1. **No Multicolinearity:**

There should not be linear relationship between independent variables. All independent variables should be independent of each other.

1. **Normality of Residual:**

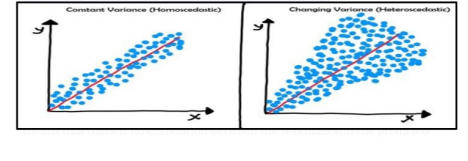
Underlying residuals are normally distributed. Residuals should be within the same range. (It uses density plot and Hypothesis testing for this)



If the residuals are not normally distributed, their randomness is lost, which implies that the model is not able to explain the relation in the data.

1. **Homoscedascity of residual or no Heteroscedascity:**

It is assumed that the residual terms have the same (but unknown) variance, σ2 This assumption is also known as the assumption of homogeneity or homoscedasticity.



1. **What if these assumptions get violated?**

'''Our Model might get overfitted or underfitted'''

1. **What is the assumption of homoscedasticity?**

**7.What is the assumption of normality?**

1. Normality of residual means the mean value of the residual should be nearly equal to 0'
2. All the residuals(error) should be normally distributed.
3. Underlying residuals are normally distributed. Residuals within the same range. (It uses density plot and Hypothesis testing for this) If the residuals are not normally distributed, their randomness is lost, which implies that the model is not able to explain the relation in the data.

**8. How to prevent heteroscedasticity?**

1. '''Detect and Handle outliers, because this linear regression algorithm is outlier sensitive