ANALYZING THE TITANIC DATA SET

QN. What is the difference between logistic regression and linear regression?

Difference 1: Variable Type of Response

A linear regression model used when the response variable takes on a continuous value such as:

- Price

- Height

- Age

- Distance

Conversely, a logistic regression model is used when the response variable takes on a categorical value such as:

- Yes or No

- Male or Female

- Win or Not Win

Difference 2: Equation Used

Linear regression uses the following equation to summarize the relationship between the predictor variable(s) and the response variable:

$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + ... + \beta_p X_p$

where:

- Y: The response variable

- X_j: The jth predictor variable

- β_j: The average effect on Y of a one unit increase in X_j, holding all other predictors fixed

Conversely, logistic regression uses the following equation:

$p(X) = e^{\beta_0} + \beta_1 X_1 + \beta_2 X_2 + ... + \beta_p X_p / (1 + e^{\beta_0} + \beta_1 X_1 + \beta_2 X_2 + ... + \beta_p X_p)$

This equation is used to predict the probability that an individual observation falls into a certain category.

Difference 3: Method Used to Fit Equation

Linear regression uses a method known as **ordinary least squares** to find the best fitting regression equation.

Conversely, logistic regression uses a method known as **maximum likelihood estimation** to find the best fitting regression equation.

Difference 4: Output to Predict

Linear regression predicts a continuous value as the output. For example:

- Price (150,199, $400.)

- Height (14 inches, 2 feet, 94.32 centimeters.)

- Age (2 months, 6 years, 41.5 years.)

- Distance (1.23 miles, 4.5 kilometers.)

Conversely, logistic regression predicts probabilities as the output. For example:

- 40.3% chance of getting accepted to a university.

- 93.2% chance of winning a game.

- 34.2% chance of a law getting passed.