Linear Probability Model

- A linear probability model (LPM) is a dummy variable regression where the dependent variable is linear in the parameters.

$$D = \beta_0 + \beta_1 X + \epsilon$$

- Conditional expectation of the dependent variable equals the conditional success probability.

$$E(D|X) = 1 \cdot P(D = 1|X) + 0 \cdot P(D = 0|X) = P(D = 1|X) = \beta_0 + \beta_1 X$$

- Errors is a binary random variable and heteroskedastic.

$$\epsilon = D - \beta_0 - \beta_1 X = \begin{cases} -\beta_0 - \beta_1 X & \text{if } D = 0 \\ 1 - \beta_0 - \beta_1 X & \text{if } D = 1 \end{cases} \quad Var(\epsilon|X) = P(1-P) = (\beta_0 + \beta_1 X)(1 - \beta_0 - \beta_1 X)$$

Logistic Regression

- The logistic regression model is a non-linear model that uses the logistic function to model the probability of a binary outcome.
- The logistic function is defined as:

$$P(D=1|X) = F(Z) = \frac{1}{1+e^{-Z}}$$

  where Z is a linear combination of the independent variables.
- The logit model estimates the coefficients using maximum likelihood estimation (MLE).
- The logit model can be interpreted in terms of odds ratios, i.e. the ratio of the probability of success to the probability of failure.

$$\frac{P(D=1|X)}{P(D=0|X)} = \frac{P}{1-P} = e^{Z} \Leftrightarrow \ln\left(\frac{P}{1-P}\right) = Z = \beta_0 + \beta_1 X + \epsilon$$