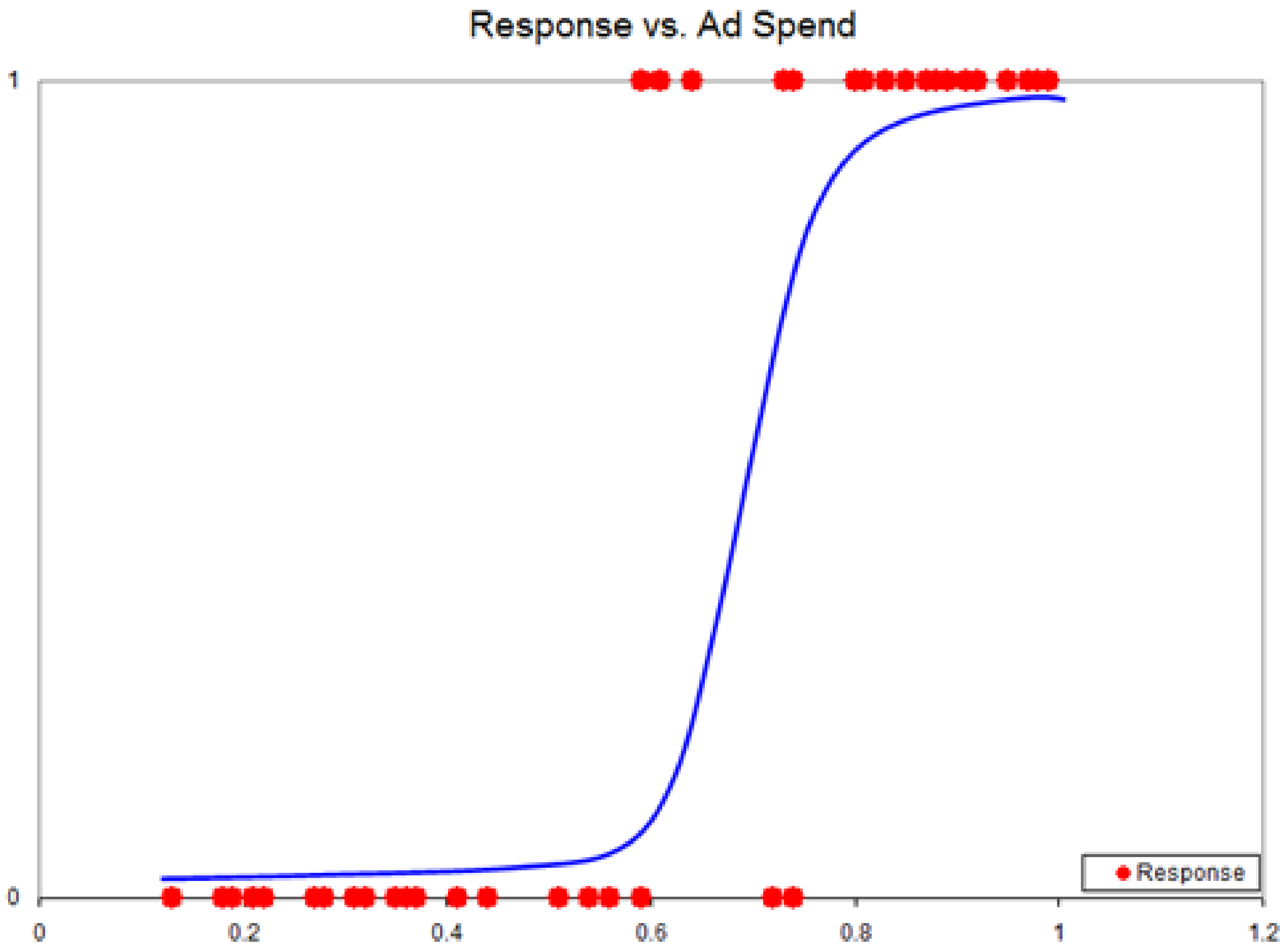
Machine Learning INF2008

Lecture 04: Logistic Regression

Donny Soh

Singapore Institute of Technology

Motivation for Logistic Regression



Probability versus Odds

$$probability = \frac{outcome}{probabilities}$$

$$prob(heads) = \frac{1}{2}$$

$$prob(dice = 3) = \frac{1}{6}$$

- What would the probability of a dice turning up the number 4 or 5 in a single toss be?
- What would the probability of a dice turning up the number 4 and 5 in a single toss be?
- What is the probability of winning any number in the Singapore Pools 4D?

$$odds = \frac{P}{1 - P}$$

$$odds(heads) = \frac{{}^{1}/_{2}}{1 - {}^{1}/_{2}}$$

$$odds(heads) = \frac{{}^{1}/_{2}}{{}^{1}/_{2}}$$

$$odds(heads) = 1$$

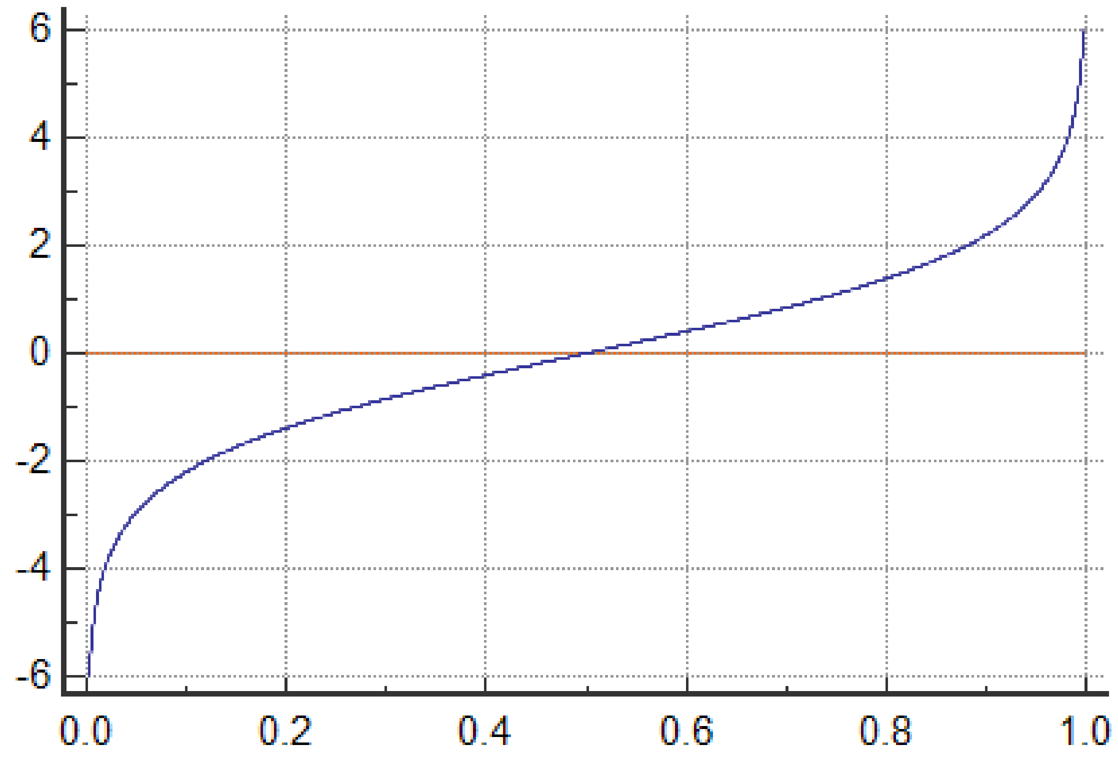
- What is the odds of a dice turning up the number 3?
- What does it mean when the odds is equal to one?
- What is the min value of odds?
- What is the max value of odds?

Formula for Logistic Regression

log of odds (logit)

$$log_e \frac{p}{1-p} = \beta_0 + \beta_1 x_1 + \beta_2 x_2$$

Linear regression



$$y = log_e \frac{p}{1-p}$$

$$e^y = \frac{p}{1-p}$$

$$e^y(1-p) = p$$

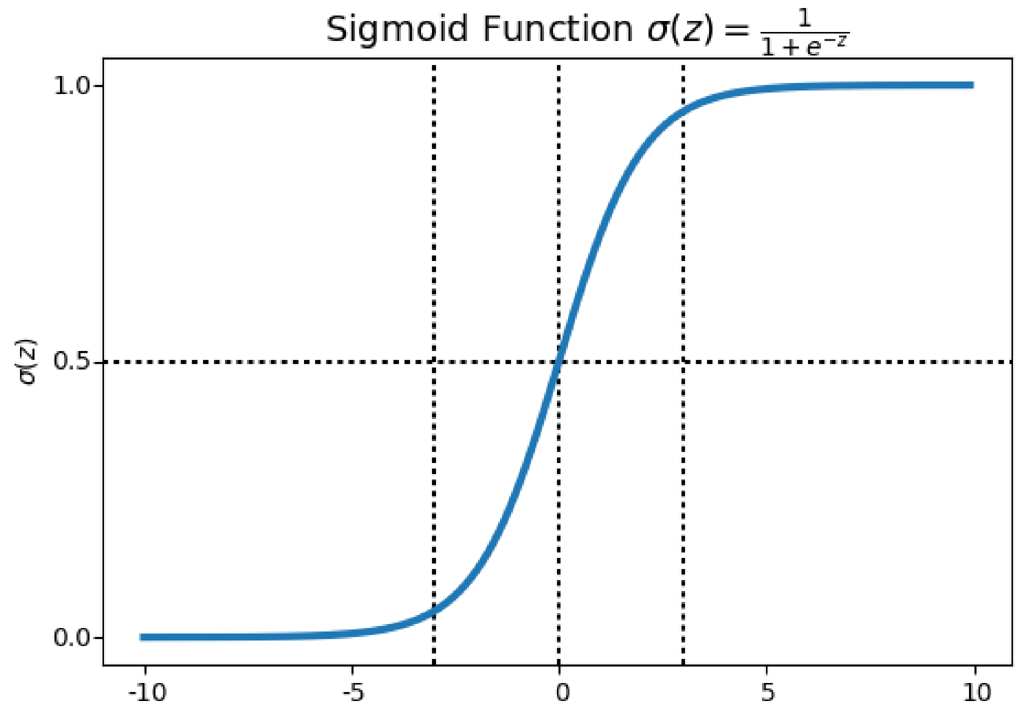
$$e^y - e^y * p = p$$

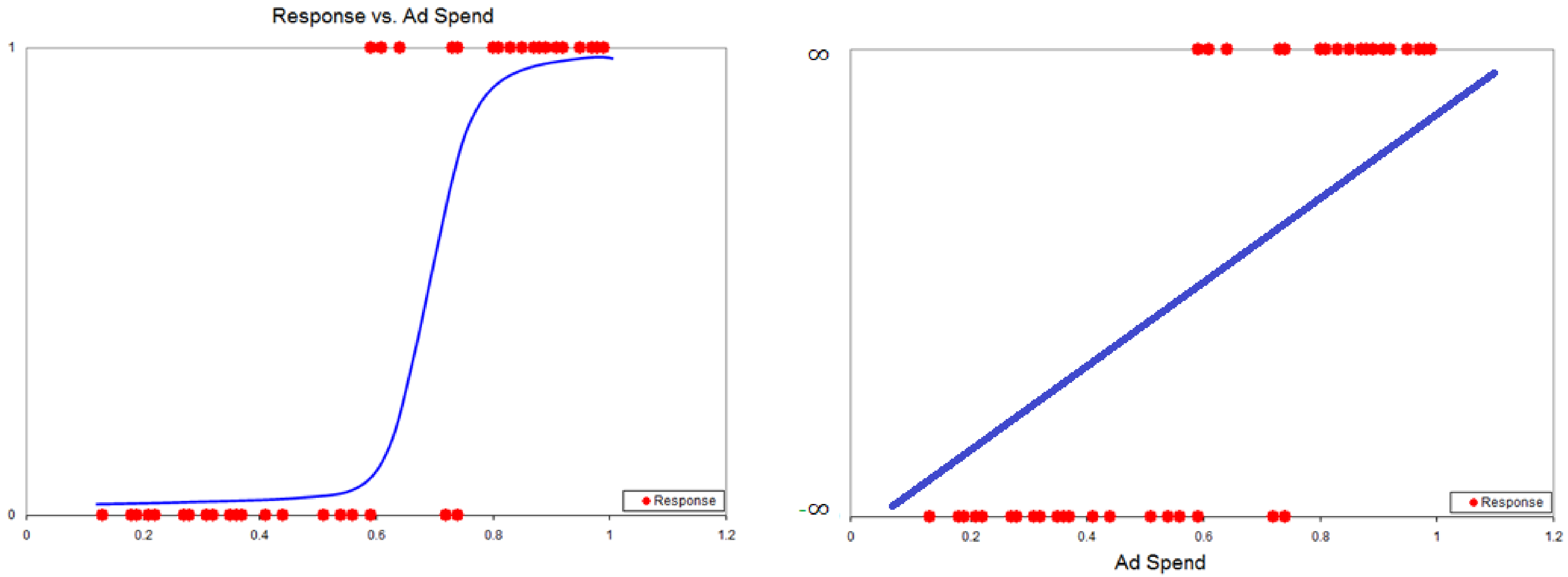
$$e^y = p + e^y * p$$

$$e^y = p(1+e^y)$$

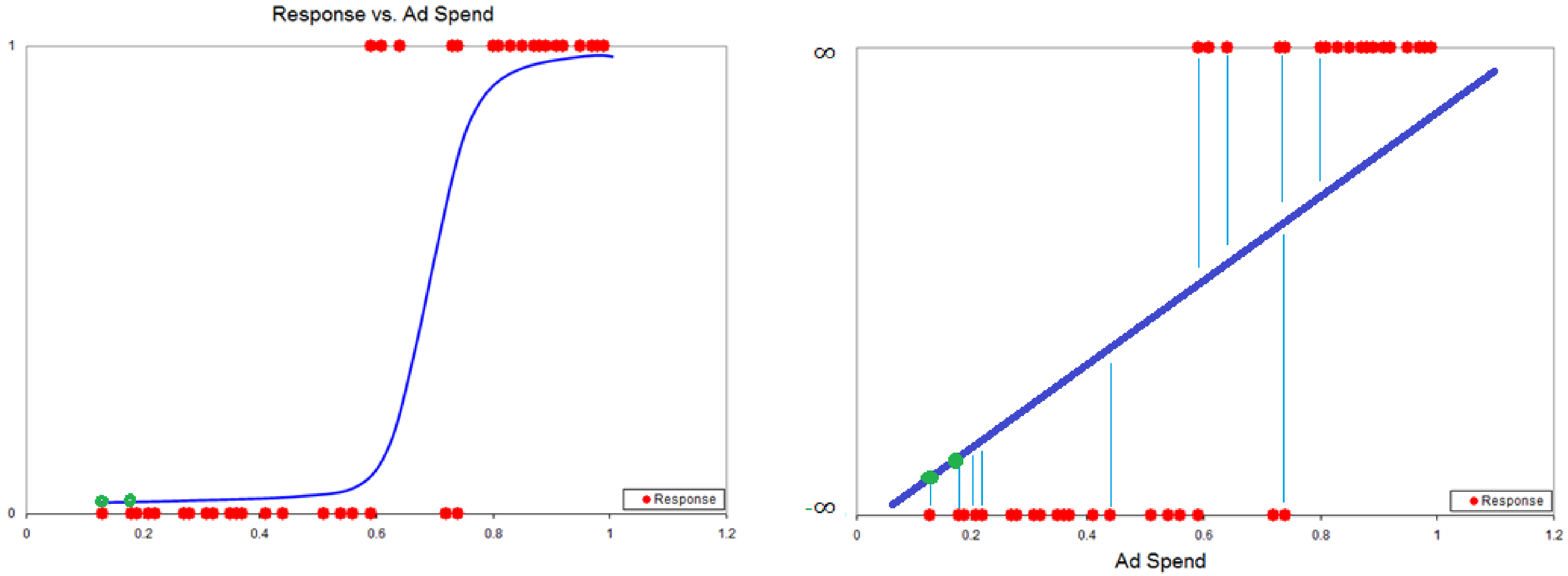
$$p = \frac{e^y}{(1+e^y)}$$

$$p = \frac{1}{(1+e^{-y})}$$

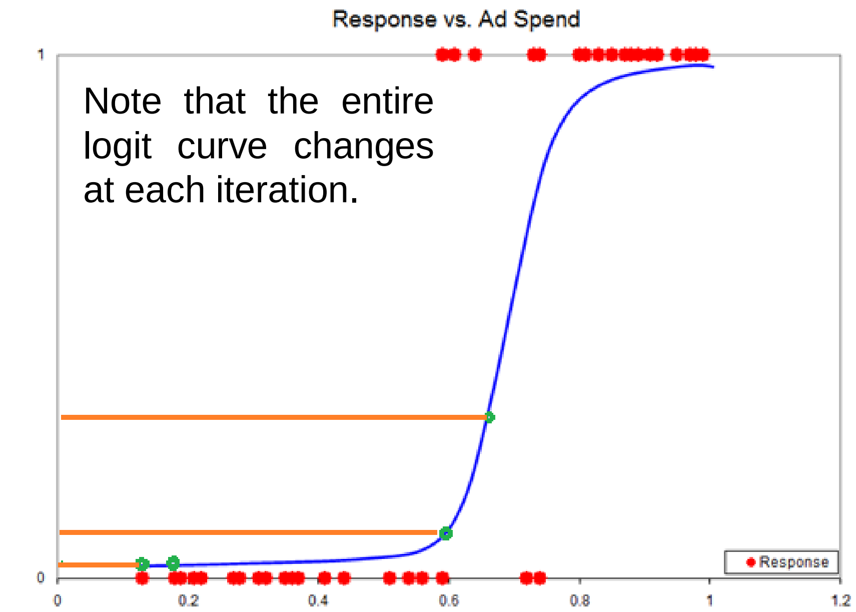


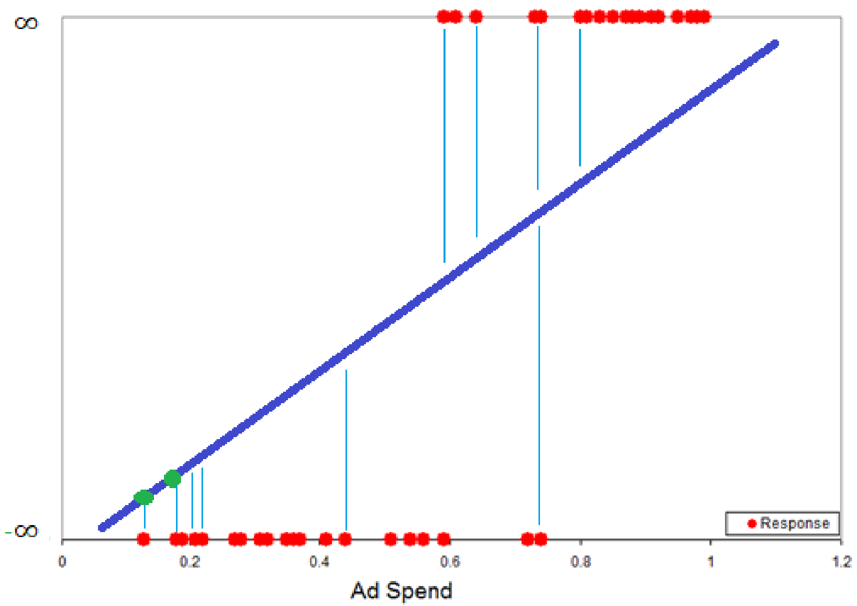


Step 1. Assume a straight line on the graph of log-odds versus x. In our running example, it will be a graph of
log odds(ad clicked) versus ad spend

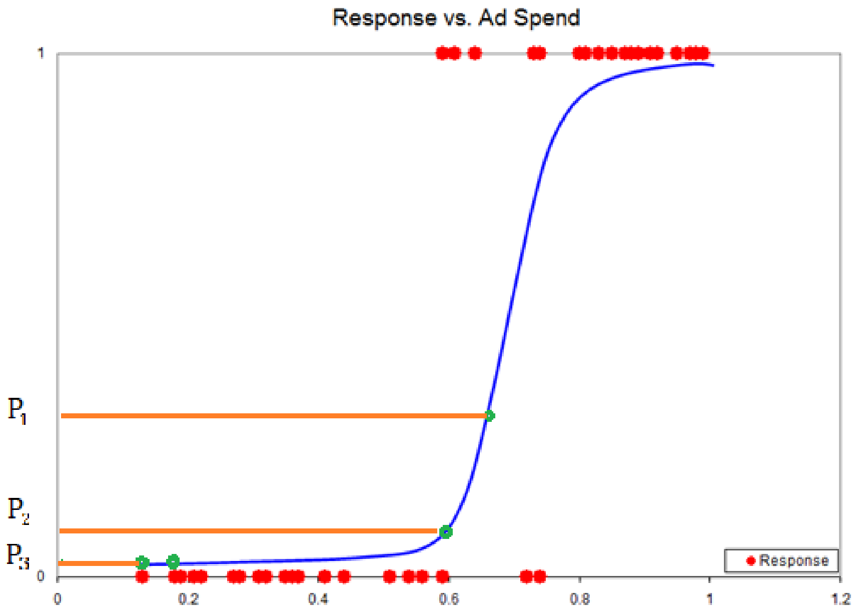


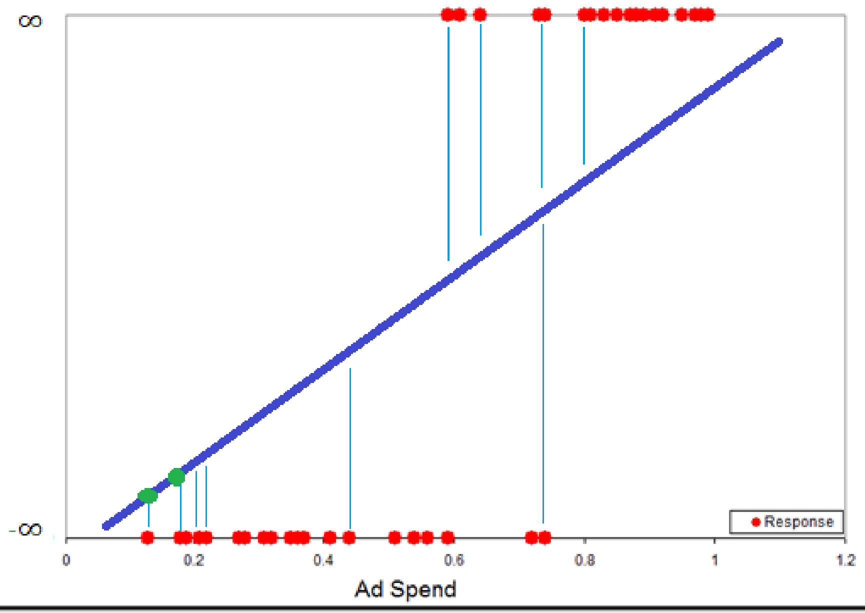
Step 2. Project the x values of ad spend on this straight line





Step 3. Find the log-odds and calculate the probability from the log-odds





Step 4. From the probability, one can then calculate what is the likelihood of the data appearing, given the model for each single point. The probabilities of all the points should be multiplied together. This gives the likelihood of the data given the coefficient