Logistic Regression

Course Starts at 6:35 PM

Logistic Regression

$$\pi(x) = \frac{e^{\alpha + \beta_1 X_1 + \beta_2 X_2 + \ldots + \beta_p X_p}}{1 + e^{\alpha + \beta_1 X_1 + \beta_2 X_2 + \ldots + \beta_p X_p}}$$

Logistic Regression

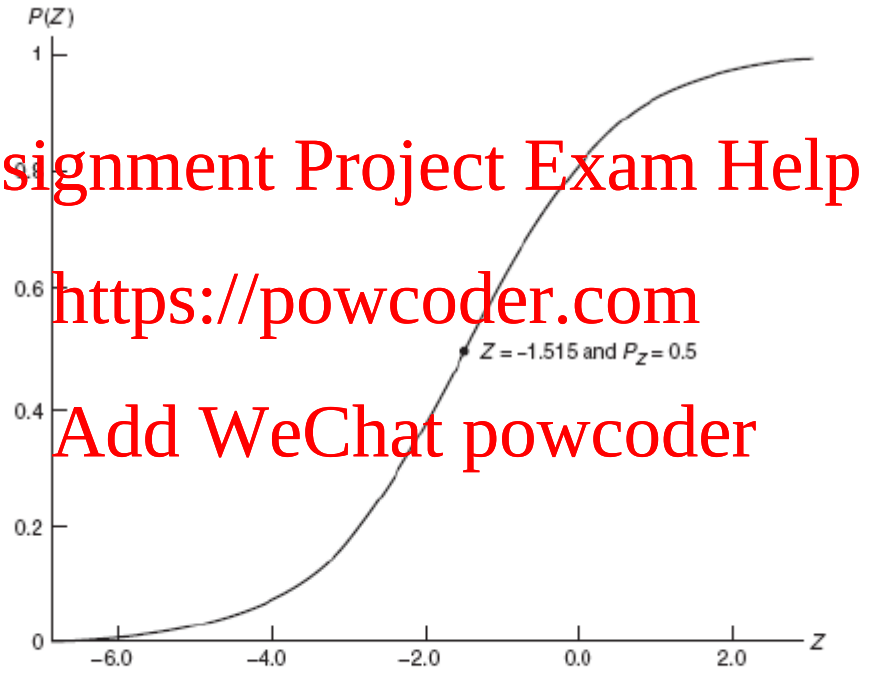

Figure 12.1: Logistic Function for the Depression Data Set

This is a Logistic Function – Hence Logistic Regression

Logistic Regression – Model
Odds are multiplicative

- ln (odds) = $\alpha + \beta_1 X_1 + \beta_2 X_2 + \ldots + \beta_p X_p$
- Odds = $e^{\alpha + \beta_1 X_1 + \beta_2 X_2 + \ldots + \beta_p X_p}$
  - $= (e^{\alpha})(e^{\beta_1 X_1})(\ldots)(e^{\beta_p X_p})$
- Or: Odds = constant * exp (constant * X_1)

  . . .

  * exp (constant * X_p)

Logistic Regression – Model Log (Odds) are additive

- $\ln(\text{odds}) = \alpha + \beta_1 X_1 + \beta_2 X_2 + \ldots + \beta_p X_p$

- Or: ln (Odds) = constant + (constant * X_1)

. . .

+ (constant * X_p)

Logistic Regression – Model Flexibility

- X (independent) variables can be continuous or categorical
- Interactions can be incorporated
- Coefficients are estimated by maximum likelihood
- Most computer programs implicitly use prior probabilities estimated from sample.

Logistic Regression – Model
Generalized Linear Model (GLM)

- Logistic regression is an example of the GLM
- Define Y = outcome = 1 (event) or 0 (not)
- E (Y|X's) = μ = P(1|X's)
- Find a function g(μ), called the *link* function, such that:

$$g(\mu) = \text{linear function of the X's}$$

- This is called the GLM
- Here we take g(μ) = ln (odds) = logit function

Logistic Regression – Model Estimation

- Model is: $g(\mu) = \alpha + \beta_1 X_1 + \beta_2 X_2 + \ldots + \beta_p X_p$
- Need to estimate: $\alpha, \beta_1, \beta_2, \ldots \beta_p$
- Use an interative process called Iterative Weighted Least Squares:
  1. Start with initial estimates of parameters
  2. Evaluate the score equations (derivative of log-likelihood = 0)
  3. Solve the score equations and get new estimates of parameters
  4. Repeat until convergence.

Logistic Regression – Model
Example: Depression Data Set
Adjusted Risk Ratio

- RR = P(Y=1|X=1)/P(Y=1|X=0)
- $P(Y=1|X) = e^{LC} / (1 + e^{LC})$, $LC = A + B_1 X_1 + \ldots$
- Example: Depression, X = sex = 1 if F, age = 30, income = 10 ($10K/year)
- Find adjusted RR for F vs. M
- LC = -0.676 – 0.021 Age – 0.037 Income + 0.929 Sex