Peter Ye
Assignment 1
ADSP31010 IP01

1. For each of the situations below, identify the response and explanatory variables, variable types, and the generalized linear model that is well-suited to model the data. Make sure to justify your answer. Write down the linear predictor as well as the link function

- **The effect of age, sex, height, daily food intake and minutes of daily exercise on a person's weight.**

  Response Variable: *weight* (numerical, continuous)

  Explanatory Variable: *age* (numerical, continuous/discrete), *sex* (categorical, binary), *height* (numerical, continuous), *daily food intake* (numerical, continuous), *minutes of daily exercise* (numerical, continuous)

  GLM: A normal linear model because the response variable *weight* is continuous and l likely to be normal distributed

  Linear Predictors: $\beta_0 + \beta_1 * age + \beta_2 * sex + \beta_3 * height + \beta_4 * daily\ food\ intake + \beta_5 * daily\ exercise$

  Link Function: $g(\mu_i) = \mu_i$

- **The effect of sex, age, GPA, major, prior years of work experience, and prior income levels on whether a full-time graduate student finds employment upon graduation**

  Response Variable: employment status upon graduation (categorical, binary)

  Explanatory Variable: *sex* (categorical, binary), *age* (numerical, continuous/discrete), *GPA* (numerical, continuous), *major* (categorical, nominal), *prior years of work experience* (numerical, continuous), *income level* (categorical, nominal)

  GLM: A logistic regression model because the response variable is categorical and binary

  Linear Predictors: $\beta_0 + \beta_1 * age + \beta_2 * \text{sex} + \beta_3 * GPA + \beta_4 * major + \beta_5 * prior\ years\ of\ work\ + \beta_6 * income\ levels$

  Link Function: $g(\mu_i) = \log\left(\frac{\mu_i}{1-\mu_i}\right)$

- **The number of mortgage loan defaults in a given year by different counties across the United States. For each household/borrower information on income, loan interest rate, age, debt, loan to value at origination are available.**

Peter Ye
Assignment 1
ADSP31010 IP01

Response Variable: the number of mortgage loan defaults (numerical, discrete)

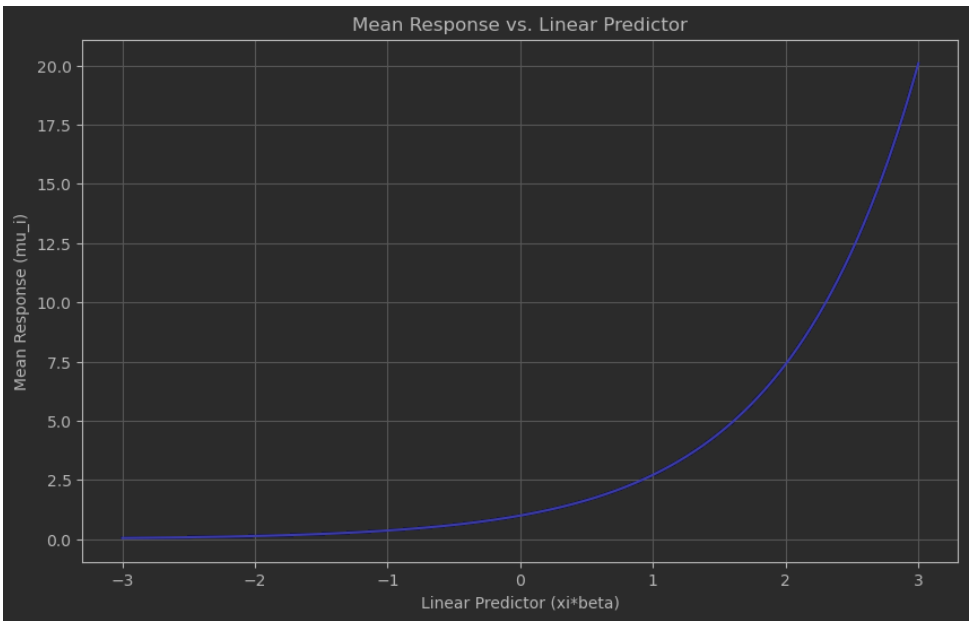
Explanatory Variable: *income* (numerical, continuous), *loan interest rate* (numerical, continuous), *age* (numerical, continuous/discrete), *debt* (numerical, continuous), *loan to value at origination* (numerical, continuous).

GLM: A binomial regression model because the response variable is numerical and discrete making sure it's greater than 0.

Linear Predictors: $\beta_0 + \beta_1 * income + \beta_2 * load\ interest\ rate + \beta_3 * age + \beta_4 * debt + \beta_5 * load\ to\ value\ at\ origination$

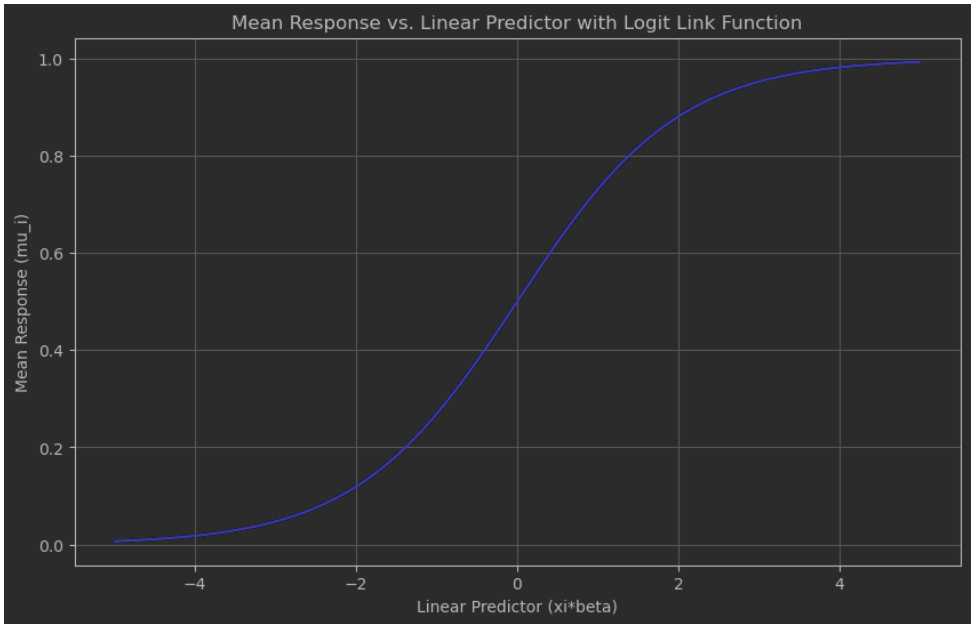
Link Function: $g(\mu_i) = \log(\mu_i)$

2. Assume that the linear predictor, $\boldsymbol{x_i^T \beta}$, takes the values between -3 and 3 equally spaced by 0.01. If the link function is the natural logarithm, $\boldsymbol{log(\mu_i) = x_i^T \beta}$, then compute and plot the mean response, $\boldsymbol{\mu_i}$, against the linear predictor, $\boldsymbol{x_i^T \beta}$.



Peter Ye
Assignment 1
ADSP31010 IP01

3. Assume that the linear predictor, $x_i^T\beta$, takes the values between -5 and 5 equally spaced by 0.01. If the link function is logit, $\log\left(\frac{\mu_i}{1-\mu_i}\right) = x_i^T\beta$, then compute and plot the mean response, μ_i, against the linear predictor, $x_i^T\beta$.



4. Assume that the linear predictor, $x_i^T\beta$, takes the values between 0.5 and 5 equally spaced by 0.01. If the link function is negative inverse, $-\mu_i^{-1} = x_i^T\beta$, then compute and plot the mean response, μ_i, against the linear predictor, $x_i^T\beta$.