1. Let us consider the dataset retinopathy.txt:

```
    RET SM  DIAB        GH        BP
1     0  1  7.04  8.472001  91.60000
2     2  1 16.31 11.276756  77.75000
3     2  0 24.46  7.910171  95.80000
4     0  1  6.57  7.065646  74.20000
5     2  1 27.94  9.970089  79.50000
6     1  1 22.21  6.752969  78.40000
.

A data frame with 613 observations on the following 5 variables.
RET=0: no retinopathy, RET=1 nonproliferative retinopathy,
RET=2 advanced retinopathy or blind
SM=1: smoker, SM=0: non-smoker
DIAB - diabetes duration in years
GH - glycosylated hemoglobin measured in percent
BP - diastolic blood pressure in mmHg
```

Let us denote the variables as Y = RET and X_1 = DIAB, X_2 = GH, X_3 = BP, X_4 = SM.

(a) Consider the multinomial logit model ($k = 2, 3$):

$$\log\left(\frac{\theta_{ik}}{\theta_{i1}}\right) = \beta_{0k} + \beta_{1k}x_{i1}.$$

Calculate the maximum likelihood estimate for the probability $P(Y_{i_*} = "2"$ = advanced retinopathy or blind), when $x_{i_*1} = 20$.

(2 points)

(b) Consider the cumulative proportional odds logit model

$$\text{logit}\,(P(Y_i \leq k)) = \beta_{0k} - \beta_1 x_{i1}, \qquad k = 1, 2.$$

Calculate the maximum likelihood estimate for the probability $P(Y_{i_*} = "2"$ = advanced retinopathy or blind), when $x_{i_*1} = 20$.

(2 points)

(c) Consider modeling the response variable Y=RET by following two different models:

$$\begin{aligned} \mathcal{M}_1: &\quad \text{logit}\,(P(Y_i \leq k)) = \beta_{0k} - \beta_1 x_{i1}, \\ \mathcal{M}_{1|2|3|4}: &\quad \text{logit}\,(P(Y_i \leq k)) = \beta_{0k} - (\beta_1 x_{i1} + \beta_2 x_{i2} + \beta_3 x_{i3} + \beta_4 x_{i4}). \end{aligned}$$

Calculate the mean square error value $\text{MSE} = \frac{\sum_{i=1}^{n}(\mathbf{y}_i - \hat{\boldsymbol{\theta}}_i)'(\mathbf{y}_i - \hat{\boldsymbol{\theta}}_i)}{n}$ for these models. Which model fits best to the data?

(2 points)

2. Based on the data in file NitrogenYield.txt, model how the nitrogen content in fertilization $X =$ Nitrogen affects the amount of yield measured in variable $Y =$ Yield.

```
   Nitrogen  Yield
1        10  28.08
2        10  30.92
3        10  30.71
4        20  37.86
.
59      200 105.55
60      200  92.17

Description: Pounds of Nitrogen (lbs/acre) and yield (bushels) for plots.
X=Nitrogen/acre  (lbs), Y=Yield (bushels)
Source: P.R. Johnson (1953). "Alternative Functions for Analyzing
a Fertilizer-Yield Relationship", Journal of Farm Economics, Vol. 35, #4, pp 519-529.
```

Let us assume the normality $Y_i \sim N(\mu_i, \sigma^2)$.

(a) Model the expected value μ_i by the second degree polynomial model

$$\mu_i = \beta_0 + \beta_1 x_i + \beta_2 x_i^2.$$

Calculate the maximum likelihood estimate for the parameter β_2.

(1 point)

(b) Model the expected value μ_i by the exponential model

$$\mu_i = e^{\beta_0} x_i^{\beta_1}.$$

Calculate the maximum likelihood estimate for the expected value μ_{i_*}, when $x_{i_*} = 150$.

(1 point)

(c) Model the expected value μ_i by the asymptotic regression, SSasymp, model

$$\mu_i = \beta_0 + (\beta_1 - \beta_0)e^{(-e^{\beta_2} x_i)},$$

where $\beta_0, \beta_1, \beta_2$ are unknown parameters. Calculate the maximum likelihood estimate for the parameter β_0.

(1 point)

(d) Model the expected value μ_i by the Michaelis-Menten, SSmicmen, model

$$\mu_i = \frac{\beta_1 x_i}{\beta_0 + x_i},$$

where β_0, β_1 are unknown parameters. Calculate the maximum likelihood estimate for the expected value μ_{i_*}, when $x_{i_*} = 150$.

(1 point)

(e) Consider again the asymptotic regression, SSasymp, model

$$\mu_i = \beta_0 + (\beta_1 - \beta_0)e^{(-e^{\beta_2} x_i)}.$$

Create 80% prediction interval for new observation y_f, when $x_f = 150$.

(2 points)

3. (a) Let the random variable Y_i be defined on ordinal scale with m distinctive possible outcomes. Let the possible outcomes have natural order "1" < "2" < "3". Consider cumulative proportional odds logit model

$$\log\left(\frac{P(Y_i \leq k)}{1 - P(Y_i \leq k)}\right) = \text{logit}\left(P(Y_i \leq k)\right) = \beta_{0k} + \beta_1 x_{i1}, \qquad k = 1, 2.$$

Solve the probabilities $P(Y_i = 1), P(Y_i = 2), P(Y_i = 3)$ as functions of parameters β_{0k}, β_1.

(3 points)

(b) In generalized linear models, the likelihood equations can written in form

$$\frac{\partial l(\boldsymbol{\beta}, \phi)}{\partial \beta_j} = \sum_{i=1}^{n} \frac{(y_i - \mu_i)}{\text{Var}(Y_i)} x_{ij} \left(\frac{\partial \mu_i}{\partial \eta_i}\right) = 0, \qquad j = 0, 1, 2 \ldots p.$$

Consider now the simple logit model with

$$Y_i \sim Ber(\mu_i),$$
$$\text{logit}(\mu_i) = \eta_i = \beta_0.$$

What kind of more simplified form the likelihood equations have in this case? That is, what form $\frac{\partial l(\beta_0)}{\partial \beta_0}$ has in the simple logit model? By using the likelihood equations, find the maximum likelihood estimator $\hat{\beta}_0$.

(3 points)