1. Consider the GLM with
$$Y_i \sim \text{Exp}(\mu_i), \text{with } \mu_i = \exp(\boldsymbol{x}_i^T\boldsymbol{\beta}).$$
where $\text{Exp}(\mu)$ is the exponential distribution with $f(y) = \mu e^{-y\mu}$, so that $\mathsf{E}[Y] = \mu$ and $\text{var}(Y) = \mu^2$.
   (a) What is the canonical link function in this case?
   (b) Derive the score function and the Hessian matrix of the log-likelihood function.

2. We have seen in class that the Poisson regression is particularly useful in modeling the count of a specific event of interest. However, in real-world applications, the numbers of zeros in the sample may not be properly modeled by a Poisson distribution. That is to say, conditioning on having a positive count, a Poisson fits the data well, but that is not true for the probability of having a zero count. To address this problem, one may modified the Poisson regression as follows: Y_i is exactly zero with probability $p(\boldsymbol{X}_i)$, and Y_i is a $\text{Poisson}(\lambda(\boldsymbol{X}_i))$ random variable with probability $1 - p(\boldsymbol{X}_i)$. Recall that the Poisson density is given by
$$f(k;\lambda) = \frac{\lambda^k}{k!}e^{-\lambda}, \quad \text{for } \lambda > 0.$$
   (a) Let $p_i = p(\boldsymbol{X}_i)$ and $\lambda_i = \lambda(\boldsymbol{X}_i)$, calculate $\mathsf{P}(Y_i = 0|\boldsymbol{X}_i)$ and $\mathsf{P}(Y_i = k|\boldsymbol{X}_i)$ for $k > 0$.
   (b) Under the GLM framework, we model the probability by
$$\text{logit}(p_i) = \log\left(\frac{p_i}{1-p_i}\right) = \boldsymbol{X}^T\boldsymbol{\beta}, \quad \text{and} \quad \log(\lambda_i) = \boldsymbol{X}^T\boldsymbol{\gamma}.$$
   Write down the likelihood function of this model, given data $\{(\boldsymbol{X}_i, Y_i), i = 1\ldots, n\}$.
   (c) To find the MLE for $\boldsymbol{\beta}, \boldsymbol{\gamma}$, calculate the gradient and Hessian of the likelihood function you derived from part (b).
   (d) Describe a procedure to justify the use of this modified model instead of the naive Poisson regression.

3. The data in the table are times to death, y_i, in weeks from diagnosis and $\log_{10}$(initial white blood cell count), x_i for seventeen patients suffering from leukemia.

x_i	65	156	100	134	16	108	121	4	39
y_i	3.36	2.88	3.63	3.41	3.78	4.02	4.00	4.23	3.73
x_i	143	56	26	22	1	1	5	65	
y_i	3.85	3.97	4.51	4.54	5.00	5.00	4.72	5.00	

   (a) Plot y_i against x_i. Do the data show any trend?
   (b) A possible specification for $\mathsf{E}[Y]$ is
$$\mathsf{E}[Y_i] = \exp(\beta_1 + \beta_2 x_i)$$
   The exponential distribution is often used to describe survival times. Fit a generalized linear model for the equation of $\mathsf{E}[Y]$ above and the exponential distribution for Y using appropriate statistical software.

(c) For the model fitted in (b), compare the observed values y_i and fitted values $\widehat{y}_i = \exp(\widehat{\beta}_1 + \widehat{\beta}_2 x_i)$. Use the standardized residuals $r_i = (y_i - \widehat{y}_i)/\widehat{y}_i$ to investigate the adequacy of the model. (Note: $\widehat{y}_i$ is used as the denominator of r_i because it is an estimate of the standard deviation of Y_i.)

4. Suppose we have grouped Bernoulli trials $Z_{ij}, 1 \le i \le I, 1 \le j \le m_i$, where $Z_{ij} \overset{i.i.d.}{\sim}$ Bernoulli(π_i). Let $Y_i = \sum_{j=1}^{m_i} Z_{ij} \sim$ Binomial(m_i, π_i). Consider the two GLMs, one for Z_{ij} and one for Y_i, both with their canonical link functions. Show that the score functions have the same value for both models.

5. l_2 **regularization of GLM.** In generalized linear model, instead of maximizing the log-likelihood, we add $\|\boldsymbol{\beta}\|_2^2$ to the log-likelihood function and maximize it. Derive the non-linear equations used to solve for $\hat{\boldsymbol{\beta}}$, as well as the gradient and Hassian used in the Newton's method.

6. Derive the explicit form of the deviance for

   (a) multiple linear regression;

   (b) logistic regression; and

   (c) Poisson regression.