Homework 2

Denis Ostroushko

Problem 1

4.2 - A

Problem 2

We define a probability of a even happening for each observation i to be a random quantity $\pi_i = P(Y = 1)$.

A GLM with a log link means that we model the natural parameter $\eta_i = log(\pi_i)$ in terms of a linear combination of predictors.

Therefore, a GLM equation is given as

$$log(\pi_i) = \hat{\beta}_0 + \hat{\beta}_1 * x_1 + \dots + \hat{\beta}_p * x_p$$

Consider the case of varying just one variable x_1 by 1 unit, which can either represent the case of switching from one categorical level to the next, or increasing a continuous predictor by 1 unit.

Changing x_1 will change the probability from π_1 to π_2, and the difference of two probabilities on the logarithmic scale is given by

$$log(\pi_2) - log(\pi_1) = \hat{\beta}_0 + \hat{\beta}_1 * (x_1 + 1) + \dots + \hat{\beta}_p * x_p - \hat{\beta}_0 - \hat{\beta}_1 * x_1 - \dots - \hat{\beta}_p * x_p =>$$

$$\hat{\beta}_1 = log(\frac{\pi_2}{\pi_1})$$

Therefore,

$$\frac{\pi_2}{\pi_1} = e^{\hat{\beta}_1}$$

. Taking the ratio instead of a difference of probabilities results in the relative comparison, therefore we evaluate relative risk here.

We do not use this link function often because of the form that $\hat{\pi}(x)$ takes on. $\hat{\pi}(x) = e^{\hat{\beta}_0+\hat{\beta}_1*(x_1+1)+\cdots+\hat{\beta}_p*x_p}$ is a function that will always be greater than 0 because of the properties of exponential function, but it is not limited by 1 on the upped end. So, given the data, we can have a scenario where fitted probabilities are greater than 1, which violates axioms of probability.

Problem 3

A

We estimate a general linear logistic regression model using a logit link function. So, taking estimates from the table, we know that the software fitted a model that takes this form:

$$log(\frac{\pi(x)}{1-\pi(x)}) = -3.7771 + 0.1449 * x$$

Using logit function, we can calculate the probability of remission when LI = 8:

$$\pi(LI = 8) = \frac{e^{-3.7771+0.1449*8}}{1+e^{-3.7771+0.1449*8}} =>$$

$\hat{\pi} = 0.068$

B

In this problem we will fix $\hat{\pi}$ at 0.5 and solve for LI.

$$log(\frac{0.5/(1-0.5)}{)} = -3.7771 + 0.1449 * x =$$

$$\frac{log(\frac{0.5}{(1-0.5)}) + 3.7771}{0.1449} = x =>$$

$x = 26.0669 \approx 26$

C

The rate of change in π in the case with one predictor is approximated by $\hat{\beta} * \hat{\pi}(x) * (1 - \hat{\pi}(x))$.

We take $\hat{\beta} = 0.1449$,while $\hat{\pi}(LI = 8) = 0.068$, from part (a). So, the rate of change is 0.1449 * 0.068 * (0.932) = 0.009

Similarly, the rate of change at LI = 26 is 0.1449 * 0.5 * 0.5 = 0.036

D

Using methods from parts (a), (b), (c) we estimate the probability of remission at LI = 14 = $\hat{\pi}(14) = P(Y = 1|LI = 14) = 0.15$.

Probability of remission at LI = 28 is $\hat{\pi}(28) = 0.57$.

Thus, probability increases by 0.42 when LI increases from 14 to 28.

E

Odds ratio for a logistic regression model is given by $e^{\hat{\beta}_1}$ for a predictor x_1. This is the multiplicative change in odds ratio.

In our problem, $\hat{\beta}_1 = 0.1449$, and so the odds ratio is $e^{0.1449} = 1.16$

F

Odds ratio is a function of the model parameter $\hat{\beta}_1$. This parameter is an MLE estimates, so by the variance property odds ratio is also an MLE. We know that MLE's are asymptotically normally distributed.

Therefore, we need to do the following steps to a confidence interval for odds ratio.

1. Get a 95% confidence interval for $\hat{\beta}_1$ using 1.96 - 97.5th quantile of the the standard normal distribution and a standard error, which we take from the model output. This is a Walk confidence interval.
2. we exponentiate the lower limit of a 95% confidence interval, an odds ratio, and an upper limit.

Recall that $\hat{\beta}_1 = 0.1449$, and the standard error is 0.0593. Therefore, the 95% confidence interval is (0.029, 0.261).

Taking an exponential of all three quantities gives us quantities that we are looking for. Odds ratio is 1.16 with a (1.03, 1.3) 95% confidence interval.

Problem 4

Problem 5

Problem 6