Problem 1

UNI: xy2395 Name: Jack Yan

i) Fill out the table and give comments.

Model	Estimate of β	CI for β	Deviance	$\hat{p}(\text{dying}\|x = 0.01)$
logit	1.162	(0.806, 1.52)	0.379	0.0901
probit	0.686	(0.497, 0.876)	0.314	0.0853
c-log-log	0.747	(0.532, 0.961)	2.23	0.128

The three link functions give different model estimates and Confidence intervals for β. The model using probit link has the smallest estimate for β. The model with probit link has the smallest deviance, indicating a good model fit. The estimated probability of death for x = 0.01 is 0.09 for the logistic model, which is different from the other two models.

ii) Suppose that the dose level is in natural logarithm scale, estimate LD50 with 90% confidence interval based on the three models.

Summary:

Model	Point estimate	CI
logit	7.39	(5.51, 9.91)
probit	7.39	(5.58, 9.90)
c-log-log	8.84	(6.53, 11.98)

a) Logit Link

$$\beta_0 + \beta_1 x_0 = g(0.5) = \log\left(\frac{0.5}{1-0.5}\right) = \log 1 = 0$$

$$\hat{x}_0 = -\frac{\hat{\beta}_0}{\hat{\beta}_1} = -\frac{-2.324}{1.162} = 2.000$$

Point estimate of LD50 is: $e^{\hat{x}_0} = e^2 = 7.39$

$$\text{var}(\hat{x}_0) = \left(\frac{\partial x_0}{\partial \beta_0}\right)^2 \text{var}(\hat{\beta}_0) + \left(\frac{\partial x_0}{\partial \beta_1}\right)^2 \text{var}(\hat{\beta}_1) + 2\left(\frac{\partial x_0}{\partial \beta_0}\right)\left(\frac{\partial x_0}{\partial \beta_1}\right) \text{cov}(\hat{\beta}_0, \hat{\beta}_1)$$

$$= \frac{1}{\hat{\beta}_1^2} \text{var}(\hat{\beta}_0) + \left(\frac{\hat{\beta}_0}{\hat{\beta}_1^2}\right)^2 \text{var}(\hat{\beta}_1) + 2\left(-\frac{1}{\hat{\beta}_1}\right)\left(\frac{\hat{\beta}_0}{\hat{\beta}_1^2}\right) \text{cov}(\hat{\beta}_0, \hat{\beta}_1)$$

$$= \frac{1}{\hat{\beta}_1^2} \text{Var}(\hat{\beta}_0) + \frac{\hat{\beta}_0^2}{\hat{\beta}_1^4} \text{Var}(\hat{\beta}_1) - 2\frac{\hat{\beta}_0}{\hat{\beta}_1^3} \text{cov}(\hat{\beta}_0, \hat{\beta}_1)$$

$$= \frac{1}{1.16^2} \times 0.175 + \frac{(-2.32)^2}{1.62^4} \times 0.033 - 2 \times \frac{(-2.324)}{1.162^3} \times (-0.066)$$

$$= 0.03184$$

$$x_0 \in [x_L, x_R] = \hat{x}_0 \pm Z_{0.05}\sqrt{\text{Var}(\hat{x}_0)} = 2.0 \pm 1.645 \cdot \sqrt{0.03184} = [1.7065, 2.2935]$$

Hence, 90% confidence interval of LD50: $[e^{x_L}, e^{x_R}] = [5.51, 9.91]$

b) Probit Link

$$\beta_0+\beta_1 X_0 = g(0.5) = \Phi^{-1}(0.5) = 0$$

$$\hat{X}_0 = -\frac{\hat{\beta}_0}{\hat{\beta}_1} = -\frac{-1.38}{0.686} = 2.006$$

Point estimate of LD50 is $e^{\hat{X}_0} = e^{2.00} = 7.39$

$$var(\hat{X}_0) = \frac{1}{\hat{\beta}_1^2} Var(\hat{\beta}_0) + \frac{\hat{\beta}_0^2}{\hat{\beta}_1^4} var(\hat{\beta}_1) - 2\frac{\hat{\beta}_0}{\hat{\beta}_1^3} cov(\hat{\beta}_0, \hat{\beta}_1)$$

$$= 0.0304$$

$$X_0 \in [X_L, X_R] = \hat{X}_0 \pm Z_{0.05} \cdot \sqrt{Var(\hat{X}_0)} = 2.0 \pm 1.645\sqrt{0.0304} = [1.7197, 2.2930]$$

Hence, 90% confidence interval of LD50: $[e^{X_L}, e^{X_R}] = [5.58,\ 9.90]$

c) Complementary Log-log link

$$\beta_0+\beta_1 X_0 = g(0.5) = \log(-\log(1-0.5)) = -0.367$$

$$\hat{X}_0 = \frac{-0.367-\hat{\beta}_0}{\hat{\beta}_1} = 2.18$$

$$Var(\hat{X}_0) = \left(\frac{\partial X_0}{\partial \beta_0}\right)^2 var(\hat{\beta}_0) + \left(\frac{\partial X_0}{\partial \beta_1}\right)^2 var(\hat{\beta}_1) + 2\left(\frac{\partial X_0}{\partial \beta_0}\right)\left(\frac{\partial X_0}{\partial \beta_1}\right) cov(\hat{\beta}_0, \hat{\beta}_1)$$

$$= \frac{1}{\hat{\beta}_1^2} var(\hat{\beta}_0) + \left(\frac{0.367+\hat{\beta}_0}{\hat{\beta}_1^2}\right)^2 var(\hat{\beta}_1) + 2\frac{1}{\hat{\beta}_1} \cdot \frac{-0.367-\hat{\beta}_0}{\hat{\beta}_1^2} cov(\hat{\beta}_0, \hat{\beta}_1)$$

$$= 0.0341$$

$$X_0 \in [X_L, X_R] = \hat{X}_0 \pm Z_{0.05}\sqrt{Var(\hat{X}_0)} = [1.88,\ 2.48]$$

Hence, 90% confidence interval of LD50: $[e^{X_L}, e^{X_R}] = [6.53,\ 11.98]$

β

Point estimate of LD50 is $e^{\hat{X}_0} = e^{2.18} = 8.84$

Problem 2

i) How does the model fit the data?

The data were fit with a logistic regression model. The residual deviance of the model is 10.613 with df = 17 - 2 = 15. The p-value for the deviance is 0.736 > 0.05, showing a good fit. The Hosmer-Lemeshow goodness of fit test also shows a good fit (p-value = 0.9907, df = 8, Chi-squared statistics = 1.6111). Therefore, we fail to reject the null and conclude that the fit is good.

ii) How do you interpret the relationship between the scholarship amount and enrollment rate? What is 95% CI?

The estimated beta_1 is 0.031, and the predictor (Amount) was fitted on the scale of thousand dollars. Therefore, the log odds of enrollment would increase by 0.031 with 1,000 dollars' increase in the scholarship.
The estimated beta_0 is -1.65, which means the log odds of enrollment would be -1.65 if no scholarship were provided. However, the interpretation of beta0 may not be useful in real life because the Amount=0 is well beyond the range of the original data. The log odds for Amount=0 may not be extrapolated from the model.

95% confidence interval is a range of values that we are 95% certain that the true parameter is contained in.

$$
\begin{aligned}
95\%\ CI \text{ for } \beta_1 &= \hat{\beta}_1 \pm z_{0.975}\, se(\hat{\beta}_1) \\
&= 0.031 \pm 1.96 \times 0.0097 \\
&= (0.012,\ 0.050)
\end{aligned}
$$

$$
\begin{aligned}
95\%\ CI \text{ for } \beta_0 &= \hat{\beta}_0 \pm z_{0.975}\, se(\hat{\beta}_0) \\
&= -1.65 \pm 1.96 \times 0.421 \\
&= (-2.47,\ -0.82)
\end{aligned}
$$

iii) How much scholarship should we provide to get 40% yield rate (the percentage of admitted students who enroll?) What is the 95% CI?

$$g(40\%) = \log\left(\frac{0.4}{1-0.4}\right) = -0.405 = \beta_0 + \beta_1 x_0$$

$$\hat{x}_0 = \frac{(-0.405) - \hat{\beta}_0}{\hat{\beta}_1} = 40.13$$

We should provide 40.13k dollars to get 40% yield rate.
(point estimate).

$$Var(\hat{x}_0) = \left(\frac{\partial x_0}{\partial \beta_0}\right)^2 var(\hat{\beta}_0) + \left(\frac{\partial x_0}{\partial \beta_1}\right)^2 var(\hat{\beta}_1) + 2\left(\frac{\partial x_0}{\partial \beta_0}\right)\left(\frac{\partial x_0}{\partial \beta_1}\right) cov(\hat{\beta}_0, \hat{\beta}_1)$$

$$= \frac{1}{\hat{\beta}_1^2} var(\hat{\beta}_0) + \left(\frac{0.405 + \hat{\beta}_0}{\hat{\beta}_1^2}\right)^2 var(\hat{\beta}_1) + 2\frac{1}{\hat{\beta}_1^2} \cdot \frac{-0.405 - \hat{\beta}_0}{\hat{\beta}_1^2} cov(\hat{\beta}_0, \hat{\beta}_1)$$

$$= 23.75$$

95% CI:

$$x_0 \in [x_L, x_R] = \hat{x}_0 \pm Z_{0.025}\sqrt{Var(\hat{x}_0)} = [30.6, 49.7]$$

The 95% CI is (30.6, 49.7) thousand dollars.