Example: Dose-Response Modeling

In the data set of interest, beetles were exposed to various levels of gaseous carbon disulphide (recorded on the $\log_{10}$ scale) for 5 hours. The observed data are given by:

j	X_j	n_j	Y_j
1	1.6907	59	6
2	1.7242	60	13
3	1.7552	62	18
4	1.7842	56	28
5	1.8113	63	52
6	1.8369	59	53
7	1.8610	62	61
8	1.8839	60	60

We model the dose response relationship between mortality and CS_2 exposure.

(a) Calculate the empirical death probability for each exposure level. Also, calculate the empirical logit, probit and complementary log-log.

- Death Prob.: $\widehat{\pi}_j = Y_j / n_j$
- Logit: $logit_j = \log\{\widehat{\pi}_j/(1-\widehat{\pi}_j)\}$
- Probit: $Probit_j = \Phi^{-1}(\widehat{\pi}_j)$
- CLL: $CLL_j = \log\{-\log(1-\widehat{\pi}_j)\}$

(b) Fit linear regression models between each empirical link and dose. Which link function appears to be most appropriate?

- See SAS code
- CLL link function has the largest R^2 value, so based on R^2 we can select CLL. However other link functions also have very high R^2 values (> 0.95), so you can select link functions based on

the context of study and the interpretation of the results.

(c) Fit a model assuming a probit link.

- Model: $\Phi^{-1}(\pi_j) = \beta_0 + \beta_1 X_j$

(d) Re-fit the model with the probit link, but this time define the end-point as $I(Y_{ij} = 0)$. What do you notice about $\widehat{\beta_1}$?

- Model:

  $\Phi^{-1}(\pi_j) = \beta_0 + \beta_1 X_j$, where $\pi_j = Pr(Y_{ij} = 1|X_j)$

- New Model:

  $\Phi^{-1}(\pi_j^*) = \beta_0^* + \beta_1^* X_j$, where
  $\pi_j^* = Pr(Y_{ij} = 0|X_j)$

  Since $\Phi^{-1}(\pi_i^*) = -\Phi^{-1}(\pi_i)$,

$$\beta_1^* = -\beta_1$$

(e) Estimate the LD50 and compare the estimate to a model-free estimator. Compute a 95% confidence interval using IML.

- LD50: x value satisfying $\pi(x) = 0.5$
- Probit model: $\Phi^{-1}\{\pi(x)\} = \beta_0 + \beta_1 x$, where $\beta_0 = -\mu/\sigma$ and $\beta_1 = 1/\sigma$.

  $\widehat{\mu} = -\widehat{\beta}_0/\widehat{\beta}_1 = 1.77$ and $\widehat{\sigma} = 1/19.72 = 0.0507$. Therefore,
  $$\widehat{LD50} = \widehat{\mu} = 1.77$$
  Model free estimator: 1.78

- By delta method
  $$Var\{g(\widehat{\beta})\} \approx \frac{\partial g}{\partial \beta}' Var(\widehat{\beta}) \frac{\partial g}{\partial \beta},$$
  where $g(\beta) = -\beta_0/\beta_1 = LD50$, and
  $$\frac{\partial g}{\partial \beta} = (-1/\beta_1, \beta_0/\beta_1^2)'$$

- Using these equations, we can show (use IML)

$$\widehat{Var}\{\widehat{LD50}\} = 0.0000145$$

- 95% CI

$$\widehat{LD50} \pm 1.96\sqrt{\widehat{Var}\{\widehat{LD50}\}} =$$

$$(1.763, 1.778)$$

(f) Fit a model assuming a complementary log-log link.

- Model: $\log\{-\log(1 - \pi_j)\} = \beta_0 + \beta_1 X_j$

(g) Estimate the LD50.

- $\log\{-\log(0.5)\} = \beta_0 + \beta_1 LD50.$

$$\begin{aligned} \widehat{LD50} &= \frac{\log\{-\log(0.5)\} - \widehat{\beta}_0}{\widehat{\beta}_1} \\ &= \frac{-0.37 + 39.57}{22.04} = 1.78 \end{aligned}$$

(h) Re-fit the model with the complementary log-log link, but this time define the end-point as $I(Y_{ij} = 0)$. What do you notice about $\widehat{\beta}_1$?

- Model:

  $\log\{-\log(1 - \pi_j)\} = \beta_0 + \beta_1 X_j$, where
  $\pi_j = Pr(Y_{ij} = 1|X_j)$

- New Model:

  $\log\{-\log(1 - \pi_j^*)\} = \beta_0^* + \beta_1^* X_j$, where
  $\pi_j^* = Pr(Y_{ij} = 0|X_j) = 1 - \pi_j$

  Since $\log\{-\log(1 - \pi_j)\} \neq -\log\{-\log(\pi_j)\}$,

$$\beta_1^* \neq -\beta_1$$

(i) Re-fit the all three models using PROC GENMOD and compare them using numeric criteria. Which link function appears to be most appropriate?

	Deviance	Pearson χ^2	Likelihood
Logit	11.23	10.03	-186
Probit	10.12	9.51	-186
CLL	3.45	3.29	-182

- CLL has the smallest deviance (and the smallest Pearson χ^2 and the largest likelihood). So we can conclude that CLL is the most appropriate link function.