Jared Yu
Module 12 Discussion

1. For the binary response variable $y = 0$ or 1, its mean is probability $p = P(y = 1)$. Most often used link function is logit link, i.e., $\eta = \ln\left(\frac{p}{1-p}\right)$. Other link functions that are used in practice:

   a. Probit link: $\eta = \Phi^{-1}(p)$, where Φ represents the cumulative standard normal distribution.
   b. Log-log link: $\eta = \ln[-\ln(1 - p)]$.
      i. (*Note: There is a correction from the discussion prompt, where a negative sign is added into the first natural log. This is based on the textbook formula for the complimentary log-log. Otherwise, the function is not calculable.*)

   Plot the three link functions and comment on how they differ.

Ans:
The plot below in Figure 1 shows each of the link functions plotted against p, the probability ranging from 0 to 1. Specifically, the plot shows the logit (green), probit (red), and complimentary log-log (blue) link functions. The plot reinforces what was stated in the book regarding the first two link functions, which is that they seem to be symmetric around the origin and $p = 0.5$, while the latter link function is off-center slightly. This can be seen since in black are horizontal and vertical dotted lines at the horizontal axis of 0.5 and vertical axis of 0 respectively.

It is interesting to note that when $p < 0.5$, the log-log seems closer to the logit. While $p > 0.5$, it starts to seem that the log-log is closer to the probit. The textbook also mentions how there is not much of a noticeable difference when the sample size is small. So it would be interesting to see how from a more empirical point of view how having one link function or another can impact the resulting classification.
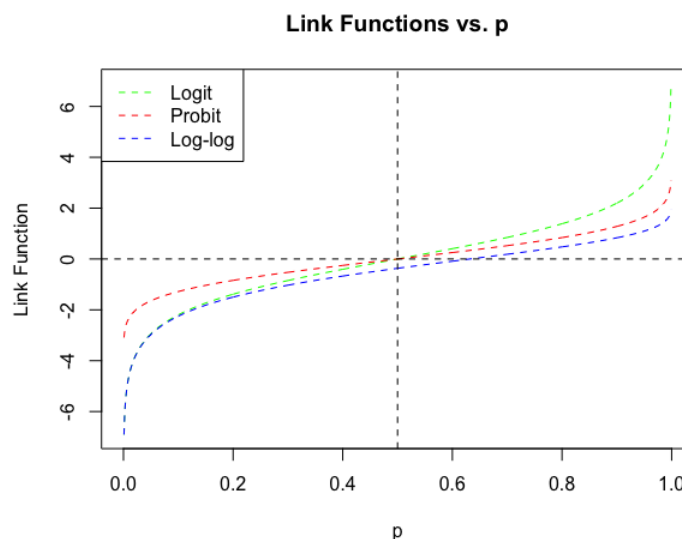

Figure 1 The above figure shows the link functions against p. The link functions include: logit (green), probit (red) and complimentary log-log (blue).

Code Appendix

```r
logit <- function(p) {
  return(log(p / (1 - p)))
}

probit <- function(p) {
  return(qnorm(p = p))
}

log_log <- function(p) {
  return(log(-log(1 - p)))
}

p_seq <- seq(0, 1, length.out = 1e3)
plot(p_seq, logit(p_seq), type = 'l',
     main = 'Link Functions vs. p',
     xlab = 'p', ylab = 'Link Function', lty = 2, col = 'green')
lines(p_seq, probit(p_seq), lty = 2, col = 'red')
lines(p_seq, log_log(p_seq), lty = 2, col = 'blue')
legend("topleft", legend = c('Logit', 'Probit', 'Log-log'),
       col = c('green', 'red', 'blue'), lty = rep(2,3))
abline(h = 0, lty = 2)
abline(v = 0.5, lty = 2)
```