Last Time: Ordinal Logistic Regression Models

- A proportional odds model is the most common way to fit an ordinal response variable.
  - One odds ratio for each explanatory variable (odds of being in lower category)
  - “Cumulative logits” = log-odds of \( Y \leq j \) versus \( Y > j \).
  - \( \beta > 0 \) implies increasing odds of being less than a given value with increasing \( x \).
  - “Inferences from fitted proportion odds models lend themselves to a general discussion of direction of response and do not have to focus on specific outcome categories.”
  - Can informally test the fit by comparing to nominal model.

Example 1: Recall the birth weights were actually a continuous response variable that we “discretized.” We can envision there being an underlying “latent” variable even with variables that are opinions (e.g., extremely unfavorable to extremely favorable). But we believe the distribution of birth weights depends on other variables, such as mother’s weight. In ordinary linear regression, we would assume we have the same distribution, but that the mean varies with the explanatory variable.

(a) Suppose our linear regression is \( \text{birth weight} - \hat{\text{hat}} = 2367 + 4.45 \times \text{mom’s weight} \) and suppose the mom weighed 125 pounds. What is the predicted birth weight? What if mom weighed 175 pounds?

\[
\begin{align*}
\text{birth} & = 2367 + 4.45(125) = 2923.35 \\
\text{birth} & = 2367 + 4.45(175) = 3145.75
\end{align*}
\]

(b) Suppose I have a normal distribution with mean 2924 and standard deviation 718, how do I find the probability below 2500? What if the mean is 3146?

\[
\begin{align*}
z_{125} & = \frac{2500 - 2924}{718} = -0.59 \ (\text{P}(z < -0.59) \approx .28) \\
z_{175} & = \frac{2500 - 3146}{718} = -0.899 \ (\text{P}(z < -0.899) \approx .194)
\end{align*}
\]

(c) If I know the predicted probability, how can I solve for the corresponding \( x \) value?

\[
\begin{align*}
x \Rightarrow z \Rightarrow \frac{\hat{z}}{\hat{\sigma}} & = \Phi(z) \Rightarrow z = \frac{x - \mu}{\sigma} \\
\Phi^{-1}(z) & = z = \frac{x - \mu}{\sigma} = -\frac{\mu}{\sigma} + \frac{1}{\sigma} x
\end{align*}
\]

This gives us a transformation or link function between the probabilities and the explanatory variable. In fact, this can be done with other continuous distribution CDFs. For example, maybe instead of following normal distributions, the latest variable follows logistic distributions.
Then the cumulative logit model holds with proportional odds. *If it is plausible to imagine an ordinary regression model with the chosen predictors describing the effects for an underlying latent variable, then it is sensible to fit the cumulative logit model with the proportional odds form.*

Additional link functions:

1. \( \ln(\pi/(1-\pi)) = \beta_0 + \beta_1 x_1 \)  
   - “logistic regression”  
   - i.e., \( \pi = \exp(\beta_0 + \beta_1 x_1)/(1 + \exp(\beta_0 + \beta_1 x_1)) \)

2. \( G^{-1}(P(Y < j)) = \beta_0 + \beta_1 x_1 \)  
   - G = logistic curve ⇒ logistic regression

3. \( F^{-1}(x) = \Phi^{-1}(\frac{e^x}{e^x + 1}) = \beta_0 + \beta_1 x_1 \)  
   - “probit regression”  
   - i.e., \( \pi = \Phi(\beta_0 + \beta_1 x_1) \)

4. \( \log(-\log(1-\pi(x))) = \beta_0 + \beta_1 x_1 \)  
   - “complementary log-log” or “Gombit”  
   - i.e., \( \pi = 1 - \exp(-\exp(\beta_0 + \beta_1 x_1)) \)  
   - asymmetric

In practice is not much difference in the fits between the logistic curve and the normal curve (especially when estimated probabilities are between .2 and .8).

**Example 2:** Sets of beetles were exposed to difference concentrations of gaseous carbon disulphide \( CS_2 \) for five hours (roughly 60 beetles at each concentration), and the proportion of insects that died after five hours of exposure were recorded. Note that dosage is measured as log concentration.

(a) Open beetles.jmp fit a logistic regression, storing the estimated probabilities.

\[
\ln \left( \frac{\hat{\pi}}{1 - \hat{\pi}} \right) = -60.74 + 34.29 \text{ (dosage)} \\
\hat{\pi} \text{ when } x = 1.691 = .059
\]

(b) At what dosage is the predicted probability of death .5 (the *median lethal dose*)?

\[
\frac{.5}{\beta} = \frac{-60.74}{34.29} = 1.77 \text{ log concentration}
\]

(c) Use JMP to fit the probit regression model (change the Link function). How do the coefficients compare? How do the estimated probabilities compare? What about the median lethal dose? How do the goodness of fit statistics compare? Is there a pattern to the residuals?

(d) How do we interpret these coefficients?

*Hint:* What is \( \text{probit}(.5) \)? \( \text{probit}(.05) \)? What is \( \hat{\pi} \) when \( x = 1.691 \)?

**Practice Problem:** Compare the fit with the “comp loglog” link.

\[
\log(-\log(1 - \hat{\pi}(x))) = -39.52 + 22.015 \text{ dosage} \\
\hat{\pi} \text{ when } x = 1.691 = .096
\]
β represents the expected number of standard deviation change in Y for a 1-unit increase in x.

π(x) = Φ((x - μ)/σ), so Φ^{-1}(π(x)) = z-score = (x - μ)/σ = (x - μ)/σ = β_0 + β_1 x