E. Collapsibility:

1. When can we ignore a variable from a model?
   a. Want parameters to have the same interpretation, and have the same estimates
   b. Depends on what parameters you are interested in.

2. Conditions:
   a. If it’s independent of all other variables
      i. ie, no interactions
      ii. Not very informative
   b. If variable is related to only one of the variables, you can collapse over it when doing inference on interactions between these variables

3. Ex., for the model $W \quad X \quad Y$
   a. ie., $\log(P[W = w, X = x, Y = y]) = \mu + \alpha^W_w + \alpha^X_x + \alpha^Y_y + \alpha^W_X wx + \alpha^X_Y xy$
   b. You can collapse over $Y$ for inference in $\alpha^W_X wx$
   c. You can collapse over $W$ for inference in $\alpha^X_Y xy$

4. Example in symbols:
   a. In full multinomial model, $P[X = x, Y = y] = \exp(\alpha^X_x + \alpha^Y_y + $
\[
\alpha_{xy} \frac{\sum_s \exp(\alpha_s W + \alpha_{sx} X)}{C} = \exp(\beta_x X + \alpha_y Y + \alpha_{xy} XY) / C
\]

for \( \beta_x X = \alpha_x X + \log(\sum_s \exp(\alpha_s W + \alpha_{sx} X)) \)

b. Definition of main effect for \( X \) changes, but interaction between \( X \) and \( Y \) stays the same.

5. MLEs exactly the same only if only a single variable is attached to \( W \).

A: 7.2, 7.3.4

F. Alternative to interaction is one-degree-of-freedom parameterization: Linear by Linear Interaction \( \alpha_{XY} = \beta_{ux vy} \)

A: 6.2-6.3

IX. Polytomous Regression:

A. Parameterization in general:

1. Take a probability associated with category \( j \) for individual \( k \)

2. Make it depend on covariates \( x_k \) through parameters \( \beta_{j} \)

3. Often force some components of \( \beta_{j} \) not to depend on \( j \)
   a. Generally first component of \( x_k \) is 1
   b. Hence first component of \( \beta_{j} \) is intercept
      i. Generally intercept depends on \( j \)
ii. Generally the other components do not.

B. Baseline-Category logits

1. \( \log(\Pr[Y_k = j]/\Pr[Y_k = 0]) = \beta_j \mathbf{x}_k \)
   a. \( (1 + \sum_{j>0} \exp(\beta_j \mathbf{x}_k)) \Pr[Y_k = 0] = 1 \)
   b. \( \Pr[Y_k = 0] = 1/(1 + \sum_{j>0} \exp(\beta_j \mathbf{x}_k)) \)

2. \( \log(\Pr[Y_k = j]/\Pr[Y_k = l]) = (\beta_j - \beta_l) \mathbf{x}_k \)

C. How do I fit this?

1. Series of separate logistic regressions conditional on sum of that category and baseline category.
   a. Let \( Z_{kj} = \begin{cases} 1 & \text{if } Y_k = j \\ 0 & \text{otherwise} \end{cases} \)
   b. \( L_j(\beta_j) = \prod_k \Pr[Y_k = 0] Z_{0j} \Pr[Y_k = j] Z_{kj} \)
   c. \( \ell_j(\beta_j) = \sum_k [Z_{0j} \log(\Pr[Y_k = 0]) + Z_{kj} \log(\Pr[Y_k = j])] = \sum_k [Z_{kj} \beta_j \mathbf{x}_k - (Z_{0j} + Z_{kj}) \log(1 + \exp(\beta_j \mathbf{x}_k))] \)
   d. \( \ell_j'(\beta_j) = \sum_k [Z_{kj} - (Z_{0j} + Z_{kj}) \exp(\beta_j \mathbf{x}_k)(1 + \exp(\beta_j \mathbf{x}_k))^{-1}] \mathbf{x}_k = \sum_k [Z_{kj} - (Z_{0j} + Z_{kj}) \pi_k] \mathbf{x}_k \)

2. All at once:
   a. \( L(\beta_1, \beta_2, \ldots, \beta_{J-1}) = \prod_k \prod_j \Pr[Y_k = j] Z_{kj} \)
   b. \( \ell \)
\[ = \sum_k \sum_j Z_{kj} \log(P[Y_k = j]Z_{kj}) \]

\[ = \sum_k \sum_{j>0} Z_{kj} \beta_j x_k + \sum_j Z_{kj} \log(P[Y_k = 0]Z_{kj}) \]

\[ = \sum_k \sum_{j>0} Z_{kj} \beta_j x_k - \sum_j Z_{kj} \log(1 + \sum_{j>0} \exp(\beta_j x_k)) \]

c. \[ \frac{d}{d\beta_j} \ell = \sum_k [Z_{kj} - \sum_l Z_{kl} \pi_{lk}] x_k \]

i. \[ \pi_{jk} = \exp(\beta_j x_k) / (1 + \sum_{l>0} \exp(\beta_l x_k)) \]

D. Cumulative logits:

1. Suppose \( \beta_j = (\theta_j, \alpha) \).

2. Suppose that \( W_k - \alpha x_k \) has CDF \( \exp(w) / (1 + \exp(w)) \)
   a. Mean 0, standard deviation 1.8138

3. Pick an increasing sequence \( \theta_j \)

4. Suppose that \( Y_k = j \) if \( W_k \in [\theta_{j-1}, \theta_j) \).

5. \( P[Y_k > j] = (1 + \exp(\theta_j + \alpha x_k))^{-1} \)

6. \( P[Y_k \leq j] = \exp(\theta_j + \alpha x_k)(1 + \exp(\theta_j + \alpha x_k))^{-1} \)

7. \( \log(P[Y_k > j] / P[Y_k \leq j]) = \theta_j + \alpha x_k : \text{cumulative logit model} \)

E. Complimentary Log-Log Link:

1. Previous analysis, with CDF \( 1 - \exp(-\exp(x)) \).
Lecture 10

a. Mean -0.577216, standard deviation 1.28255.

2. \[ \log(-\log(\Pr(Y_k > j))) = \theta_j + \alpha x_k. \]

3. Fig. 12 compares the link functions.

\begin{figure}[h]
\centering
\includegraphics[width=\textwidth]{link-comparison.png}
\caption{Link Comparison}
\end{figure}

F. Continuation logits:

1. Model log odds ratio for membership in category vs. all below it.

2. or all above it. [Mark ER] [Mark Esas]