Homework Exercise Set 2: STA 7249

1. (7.24) A multivariate generalization of the exponential dispersion family is
$$f(\mathbf{y}_i; \boldsymbol{\theta}_i, \phi) = \exp\{[\mathbf{y}_i'\boldsymbol{\theta}_i - b(\boldsymbol{\theta}_i)]/a(\phi) + c(\mathbf{y}_i, \phi)\},$$
where $\boldsymbol{\theta}_i$ is the natural parameter. Show that the multinomial variate $\mathbf{y}_i = (y_{i1}, \ldots, y_{i,J-1})'$ (with a 1 in a position if that outcome occurred, and 0 otherwise) for a single trial with parameters $\{\pi_j, j = 1, \ldots, J-1\}$ is in the $(J-1)$-parameter exponential family, with baseline-category logits as natural parameters.

2. (7.36) Suppose we express $\pi_j(\mathbf{x}) = \frac{\exp(\alpha_j + \boldsymbol{\beta}_j'\mathbf{x})}{1+\sum_{h=1}^{J-1}\exp(\alpha_h + \boldsymbol{\beta}_h'\mathbf{x})}$ as
$$\pi_j(\mathbf{x}) = \frac{\exp(\alpha_j + \boldsymbol{\beta}_j'\mathbf{x})}{\sum_{h=1}^{J}\exp(\alpha_h + \boldsymbol{\beta}_h'\mathbf{x})}.$$
Show that dividing numerator and denominator by $\exp(\alpha_J + \boldsymbol{\beta}_J'\mathbf{x})$ yields new parameters $\alpha_j^* = \alpha_j - \alpha_J$ and $\beta_j^* = \beta_j - \beta_J$ that satisfy $\alpha_J = 0$ and $\boldsymbol{\beta}_J = \mathbf{0}$. Thus, without loss of generality, $\alpha_J = 0$ and $\boldsymbol{\beta}_J = \mathbf{0}$.

3. (7.29) Is the proportional odds model a special case of a baseline-category logit model? Explain why or why not.

4. (7.35) For the cumulative probit model $\Phi^{-1}[P(Y \le j)] = \alpha_j - \boldsymbol{\beta}'\mathbf{x}$, explain why a 1-unit increase in x_i corresponds to a β_i standard deviation increase in the expected underlying latent response, controlling for other predictors.

5. (7.36) For cumulative link model $G^{-1}[P(Y \le j \mid \mathbf{x})] = \alpha_j + \boldsymbol{\beta}'\mathbf{x}$, , show that for $1 \le j < k \le J-1$, $P(Y \le k \mid \mathbf{x}) = P(Y \le j \mid \mathbf{x}^*)$ where $\mathbf{x}^*$ is obtained by increasing the ith component of $\mathbf{x}$ by $(\alpha_k - \alpha_j)/\beta_i$. Interpret.

6. (7.34) A response scale has the categories (strongly agree, mildly agree, mildly disagree, strongly disagree, don't know). One way to model such a scale uses a logit model for the probability of a don't know response and uses a separate ordinal model for the ordered categories conditional on response in one of those categories. Explain how to construct a likelihood to do this simultaneously.

7. (7.42) A cafe has four entrées: (chicken, beef, fish, vegetarian). Specify a discrete choice model for the selection of an entrée using x = gender (1 = female, 0 = male) and u = cost of entrée, which is a characteristic of the choices. Interpret the model parameters.

8. (7.1) For Table 1, let Y = belief in life after death, x_1 = gender (1 = females, 0 = males), and x_2 = race (1 = whites, 0 = blacks). Table 2 shows the fit of the model
$$\log(\pi_j/\pi_3) = \alpha_j + \beta_j^G x_1 + \beta_j^R x_2, \quad j = 1, 2,$$
with *SE* values in parentheses.

Table 1:

Race	Gender	Yes	Undecided	No
		Belief in Afterlife		
White	Female	371	49	74
	Male	250	45	71
Black	Female	64	9	15
	Male	25	5	13

Source: 1991 General Social Survey

Table 2:

Parameter	(Yes/No)	(Undecided/No)
	Belief Categories for Logit	
Intercept	0.883 (.243)	-0.758 (.361)
Gender	0.419 (.171)	0.105 (.246)
Race	0.342 (.237)	0.271 (.354)

(a) Find the prediction equation for $\log(\pi_1/\pi_2)$.

(b) Using the yes and no response categories, interpret the conditional gender effect using a 95% confidence interval for an odds ratio.

(c) Show that for white females, $\hat{\pi}_1 = \hat{P}(Y = \text{yes}) = 0.76$.

(d) Without calculating estimated probabilities, explain why the intercept estimates indicate that for black males $\hat{\pi}_1 > \hat{\pi}_3 > \hat{\pi}_2$. Likewise, use the intercept and gender estimates to show the same ordering applies for black females.

(e) Without calculating estimated probabilities, explain why the estimates in the gender and race rows indicate that $\hat{\pi}_3$ is highest for black males.

(f) For this fit, $G^2 = 0.9$. Explain why residual $df = 2$. Deleting the gender effect, $G^2 = 8.0$. Test whether opinion is independent of gender, given race. Interpret.