STA 841: Models & Methods for Categorical Data

HOMEWORK 5

Due Friday, November 10 at 11:45am

Be sure to include an R markdown file with reproducible code for questions 2 and 3.

1. **Marginal and mutual independence.** Two balanced coins are flipped independently. Let $X =$ whether the first flip resulted in a head (yes, no), $Y =$ whether the second flip resulted in a head, and $Z =$ whether both flips had the same result. Using this example, show that (marginal) independence of each pair of three variables does not imply that the variables are mutually independent.

2. **Interpreting interactions.** In a study of low birth weight (defined as an infant weighing less than 2500g), researchers are interested in the roles of maternal age (in years) and adequacy of prenatal care in predicting the outcome. They consider three variables: the outcome of low birth weight (LBW=1 if yes and 0=no), age (centered at the sample mean of 23 years so that 0 represents 23 years, -1 represents 22 years, 5 represents 28 years, etc.), and care (1=adequate prenatal care, 0=inadequate care). The investigators fit the following model to the data:

$$\text{logit}(Pr(LBW_i = 1)) = \beta_0 + \beta_1 \text{age}_i + \beta_2 \text{care}_i + \beta_3 \text{age}_i \text{care}_i$$

and obtained the estimates

$$\hat{\beta}' = (-0.52, 0.04, -0.47, -0.18)'$$

and

$$\widehat{\text{Cov}}(\hat{\beta}) = \begin{bmatrix} 0.046 & 0.002 & -0.046 & -0.002 \\ 0.002 & 0.002 & -0.002 & -0.002 \\ -0.046 & -0.002 & 0.111 & 0.004 \\ -0.002 & -0.002 & 0.004 & 0.005 \end{bmatrix}.$$

Prepare an appropriate graphical display to illustrate the associations (point and interval estimates of odds ratios) between adequacy of care, age, and low birth weight at each combination of levels of adequacy of care and maternal age (assume age takes integer values from 17-36). In addition, provide specific interpretations of each parameter estimate (including the intercept) in this model. Include reproducible code for your graphical display.

3. **Interrater Agreement for Diagnosis of Epileptic Seizures.** Consider data from a study of interrater agreement in the diagnosis of type of epileptic seizures. Two neurologists assessed the medical records of 100 patients and were asked to determine for each patient whether the patient had experienced generalized epileptic seizures, partial epileptic seizures, or no epileptic seizures. The diagnoses of both neurologists are shown below.

	Neurologist 2		
Neurologist 1	No Seizures	Partial	Generalized
No Seizures	5	4	1
Partial	7	39	9
Generalized	3	15	17

Fit the interrater agreement models discussed in class to determine the type of agreement between the two neurologists. Describe your findings, including a discussion of whether the neurologists differ in their tendency to diagnose each type of seizures overall, the type of agreement (if any) between the neurologists in their diagnoses, and any particular diagnoses that may be more challenging for the neurologists to make. Be sure to include fully reproducible code with your answer.