

BST 140.753 Take Home Final

March 10, 2005

Fine print The take home final must be completed by students individually. Please hand in no later than 3/18/05.

1. Consider the Space Shuttle data set from Homework 3. (Recall that the two 2's are treated as 1's.) If y_i is the presence or absence (0 or 1 resp.) of O-ring failure for launch i and x_i is the temperature, then consider the model $E[y_i] = \mu_i$ and $\text{logit}(\mu_i) = \beta_0 + \beta_1 x_i$.
 - a. Show that $(T_0, T_1) = (\sum_i y_i, \sum_i x_i y_i)$ are the sufficient statistics for (β_0, β_1) . Further argue that T_0 is the sufficient statistic for β_0 when β_1 is fixed.
 - b. Derive the conditional distribution of the $\{y_i\}$ given T_0 . Argue that the support of this distribution has 245,157 elements.
 - c. Plot the conditional likelihood for β_1 . Draw cut-off lines at $1/k$ for $k = 8$ and $k = 16$. Overlay the profile likelihood you calculated for Homework 3. (Normalize both so that they have maximum 1.) Compare the conditional and overall maximum likelihood estimates for β_1 .
2. Consider a GLMM for binary data using the probit link function

$$\Phi^{-1}[P(Y_{it} = 1|u_i)] = x'_{it}\beta + z'_{it}u_i$$

where Φ is the standard normal cdf and u_i is distributed as normal with mean 0 and variance Σ with pdf denoted $f(u_i; \Sigma)$.

- a. Show that the marginal mean is

$$P(Y_{it} = 1) = \int P(Z - z'_{it}u_i \leq x'_{it}\beta) f(u_i; \Sigma),$$

where Z is a standard normal variable that is distributed independently of u_i .

- b. Use the fact that $Z - z'_{it}u_i$ has a normal distribution with mean 0 and variance $1 + z'_{it}\Sigma z_{it}$ to deduce that

$$\Phi^{-1}[P(Y_{it} = 1)] = x'_{it}[1 + z'_{it}\Sigma z_{it}]^{-1/2}.$$

Hence the marginal model is also a probit model, with attenuated effects. Discuss the case where $z_{it} = 1$ and $\Sigma = \sigma^2$.

3. This problem involves using random effect models to fit capture/recapture data. Consider the survey data below where a "Present" indicates a subject was listed on that survey. If y_{ij} is a binary indicator of person i being present on survey j , consider the model $P(y_{ij} = 1) = u_i + \beta_j$. Of course, to fit this model we require n_{22} , which we treat as a parameter. Use this model to calculate a profile likelihood for n_{22} . Place $1/k$ cut-off lines at $k = 8$ and $k = 16$. Interpret your results.

First Survey	Second Survey	
	Present	Absent
Present	794	710
Absent	741	-

Table 1: Census Results

4. Consider a contingency table that cross classifies discrete random variables X , Y and Z with cell probabilities π_{ijk} . The notation, (XY, ZY) , denotes a log-linear model on the cell frequencies that has an intercept, the X , Y and Z main effects and both of the XY and ZY interactions.
- Show that the model (X, Y, Z) implies a multinomial model that has X , Y and Z as mutually independent.
 - Show that the model (XY, Z) implies a multinomial model that has X and Y as being jointly independent of Z .
 - Show that the model (XY, ZY) implies a multinomial model that has X and Z as conditionally independent given Y .
 - Show that independence of the forms a. and b. imply marginal independence of X and Z . Demonstrate via an example that conditional independence does not imply marginal independence.
 - Explore the substance abuse data below by comparing the fits of Poisson log-linear models

Alcohol Use	Cigarette Use	Marijuana Use	
		Yes	No
Yes	Yes	911	538
	No	44	456
No	Yes	3	43
	No	2	279

5. Consider the movie review rating data given below (1 - Pro, 2 - Mixed, 3 - Con). Let y_{it} be the response of reviewer t for movie i .

		Medved										
		1			2			3				
Lyons		Siskel		Ebert								
				1	2	3	1	2	3	1	2	3
1	1	15	2	0	2	0	0	8	0	2		
	2	0	3	0	2	3	0	3	2	1		
	3	1	0	1	1	2	1	1	1	2		
2	1	2	0	0	2	1	0	0	1	0		
	2	1	1	0	0	1	0	0	1	0		
	3	1	0	0	0	0	0	0	1	1		
3	1	3	1	1	0	0	0	4	1	0		
	2	0	0	1	0	0	0	1	3	0		
	3	0	0	1	2	1	0	2	2	4		

- Fit cumulative logit (proportional odds) model that includes a random intercept for movie. Include a fixed effect for reviewer. (Interpret your results.)
- Fit a baseline category logit model with random intercepts for movie and a fixed effects for reviewer. (Interpret your results.)