

STAT 770, Spring 2017: Homework 6

1. Analyze the data of Table 11.14.
 - (a) Compare the difference in marginal proportions using a 95% CI.
 - (b) Perform McNemar's test and interpret results.
 - (c) Obtain and interpret inference from fitting a conditional logistic regression model. What do the latent α_i represent in this model?
 - (d) Carefully explain the difference in the types of inference obtained in parts (a) & (b) versus (c).

2. The data for Problem 11.12 is from a pair of movie reviewers that used to have their own show (*At the Movies with Siskel and Ebert*). In their show they would colorfully discuss and rate movies.
 - (a) Test the null hypothesis of symmetry. What does symmetry imply for Siskel and Ebert?
 - (b) Summarize these data using the kappa statistic. Also compute the raw probability of agreement $\sum_{i=1}^3 \hat{\pi}_{ii}$ and interpret. How often do they agree?
 - (c) *Extra Credit:* Perform and interpret the Stuart-Maxwell test for marginal homogeneity. After installing the `coin` package, this can be carried out in R via

```
library(coin)
review=c("con","mixed","pro")
movies=as.table(matrix(c(24,8,10,8,13,9,13,11,64),nrow=3,
  dimnames=list(Siskel=review,Ebert=review)))
mh_test(movies)
```

Alternatively you can carry out Bhapkar's test; sample SAS code provided on course website.

3. Consider the respiratory illness data of Section 12.4.4 (p. 476). In class, we analyzed these data using a Markov model. Recall that the data are four repeated measurements $\mathbf{Y}_i = (Y_{i1}, Y_{i2}, Y_{i3}, Y_{i4})$ on whether a child had a respiratory illness $Y_{ij} = 1$ or not $Y_{ij} = 0$ during the year. We will consider the age of the child $t_1 = 7, t_2 = 8, t_3 = 9, t_4 = 10$ years and whether the mom smoked at baseline, $s_i = 0, 1$ for no/yes. You will analyze these data marginally using GEE, then conditionally using a logistic normal model.

- (a) Consider the marginal model with time treated as continuous

$$P(Y_{ij} = 1) = \beta_0 + \beta_1 t_j + \beta_s s_i.$$

Fit the model using `type=ar` and `type=un`: AR(1) and unstructured covariance matrices. Which working correlation $\mathbf{R}(\boldsymbol{\alpha})$ has the lower QIC?

- (b) Now consider time as categorical

$$P(Y_{ij} = 1) = \beta_0 + \beta_7 I\{t_j = 7\} + \beta_8 I\{t_j = 8\} + \beta_9 I\{t_j = 9\} + \beta_s s_i.$$

Just use `type=un`. Does this model have lower QIC than in part 1?

- (c) It appears that the time effects for years 7, 8, and 9 are all similar but different to year 10. Fit a model that only changes the log-odds of illness for year 10

$$P(Y_{ij} = 1) = \beta_0 + \beta_{10} I\{t_j = 10\} + \beta_s s_i.$$

Does this model have lower QIC than in part 2? Interpret this marginal model in terms of odds ratios. Leave smoking in the model even though it is not significant. Note that a *one-sided* test for maternal smoking is *almost* significant!

- (d) Now let's fit the logistic normal model to these data

$$P(Y_{ij} = 1) = \beta_0 + \beta_{10} I\{t_j = 10\} + \beta_s s_i + u_i, \quad u_1, \dots, u_{537} \stackrel{iid}{\sim} N(0, \sigma^2).$$

If you are using SAS PROC NLMIXED, use the coefficient values from part 3 as starting values for NLMIXED (you can leave the default starting value of $\sigma = 1$). I suggest `qpoints=50`. Interpret this model in terms of odds ratios. How is interpretation different for the GLMM vs. the marginal model?

- (e) Remove the random effects from the model (remove the `random` statement as well as `u` from `eta`) and refit the model; compare AIC for the random effects model vs. the independence model. Formally test $H_0 : \sigma = 0$ vs. $H_a : \sigma > 0$ as outlined in the notes.
- (f) Go back to the slides and read over the Markov model analysis. How does the significance of maternal smoking change when using the Markov model?
- (g) *Extra Credit*: Redo part (d) using `proc mcmc`. Do your conclusions change?

Sample SAS code

```
data resp;
  infile "c:/your STAT 770 folder/resp.txt";
  input case smoke year resp;
  d7=0; d8=0; d9=0; d10=0;
  if year=7 then d7=1; if year=8 then d8=1;
  if year=9 then d9=1; if year=10 then d10=1;

* add year to class to treat year as categorical;
proc genmod data=resp descending; class case;
  model resp=smoke year / dist=bin link=logit type3;
  repeated subject=case / type=un corrw;
run;

proc genmod data=resp descending; class case;
  model resp=smoke d10 / dist=bin link=logit type3;
  repeated subject=case / type=un corrw;
run;

proc nlmixed data=resp qpoints=50;
  parms b0=-1.7 b10=-0.4 mat_smoke=0.3;
  eta=b0+b10*d10+mat_smoke*smoke+u;
  p=exp(eta)/(1+exp(eta));
  model resp ~ binary(p);
  random u ~ normal(0, sig*sig) subject=case;
run;
```