

STAT 770, Spring 2017: Homework 5

1. Analyze the data of 6.15 (p. 245) using exact logistic regression; note that SAS reports separation in the data. Specifically, fit the additive model (homogeneous association) and test whether cure is independent of delay given penicillin level at the 5% level. Use the score test with the exact p-value. Interpret the estimated odds ratio and 95% CI for delay. Repeat the analysis using a Bayesian approach suitable for small samples. In particular, to assess how delay affects odds, exponentiate the endpoints of the equal-tailed interval for the delay parameter and compare to the conditional logistic regression approach.

```
data cure;
input level$ delay$ cured died @@;
total=cured+died;
datalines;
'1/8' none 0 6 '1/8' '1.5h' 0 5
'1/4' none 3 3 '1/4' '1.5h' 0 6
'1/2' none 6 0 '1/2' '1.5h' 2 4
'1' none 5 1 '1' '1.5h' 6 0
'4' none 2 0 '4' '1.5h' 5 0
;
proc logistic data=cure; class level delay / param=ref;
model cured/total=level delay;
exact delay / estimate=both;

proc genmod; class level delay / param=ref;
model cured/total=level delay / link=logit dist=bin;
bayes coeffprior=jeffreys;
```

2. Problem 6.27.
3. Consider the kyphosis data of problem 5.8 (p. 199).
 - (a) Fit the simple logistic regression model $\log\{\pi_i/(1 - \pi_i)\} = \beta_0 + \beta_1 a_i$, where a_i is age and π_i is the probability that kyphosis is present in child i . Plot r_i vs. a_i and superimpose a loess smooth. Comment. Does the H-L test show lack-of-fit?
 - (b) Fit the kyphosis data using a generalized additive model and formally test whether a “wiggly” part is needed on top of the linear trend.
 - (c) Finally, add a quadratic term to the model and test whether it is significant.
 - (d) Going from (a) to (b) to (c) the alternative (for H-L, deviance test in generalized additive model, and test of dropping quadratic term) becomes more and more focused. What happens to the p-value when we finally look “in the right place?” This illustrates why H-L can have low power to detect lousy models.

```

data kyphosis;
input age k @@;
datalines;
  12 1  15 1  42 1  52 1  59 1  73 1  82 1  91 1  96 1 105 1 114 1 120 1 121 1 128 1
130 1 139 1 139 1 157 1  1 0  1 0  2 0  8 0 11 0 18 0 22 0 31 0 37 0 61 0
  72 0  81 0  97 0 112 0 118 0 127 0 131 0 140 0 151 0 159 0 177 0 206 0
;

proc logistic data=kyphosis;
model k(event='1')=age / lackfit;
output out=out stdreschi=r;

proc sgscatter data=out;
plot r*age / loess;

proc gam data=kyphosis;
model k(event='1')=spline(age) / dist=bin link=logit;

proc logistic data=kyphosis;
model k(event='1')=age age*age;

```

4. Consider the data of Problem 8.4 (p. 331). Table 8.17 is actually wrong, the correct data are listed below.
- Fit a baseline-category generalized logit model to these data with gender and length as predictors. Test at the 5% level whether you can drop gender with a Type 3 test.
 - In a model that includes only length, summarize how the odds of choosing fish over invertebrates changes when length is increased by one meter. Include and interpret a 95% CI.
 - Produce and interpret a plot that shows how the probabilities of (F, I, O) change with length.
 - Set the equations for the probabilities of fish or invertebrate as functions of length equal to each other, i.e. $\pi_F(L) = \pi_I(L)$ and solve for L to find the length at which the probabilities cross.

```

data alligator;
input gender$ length food$ @@;
datalines;
M 1.30 I M 1.32 F M 1.32 F M 1.40 F M 1.42 I M 1.42 F M 1.47 I M 1.47 F
M 1.50 I M 1.52 I M 1.63 I M 1.65 O M 1.65 O M 1.65 I M 1.65 F M 1.68 F
M 1.70 I M 1.73 O M 1.78 F M 1.78 O M 1.80 F M 1.85 F M 1.93 I M 1.93 F
M 1.98 I M 2.03 F M 2.03 F M 2.31 F M 2.36 F M 2.46 F M 3.25 O M 3.28 O
M 3.33 F M 3.56 F M 3.58 F M 3.66 F M 3.68 O M 3.71 F M 3.89 F F 1.24 I
F 1.30 I F 1.45 I F 1.45 O F 1.55 I F 1.60 I F 1.60 I F 1.65 F F 1.78 I
F 1.78 O F 1.80 I F 1.88 I F 2.16 F F 2.26 F F 2.31 F F 2.36 F F 2.39 F
F 2.41 F F 2.44 F F 2.56 O F 2.67 F F 2.72 I F 2.79 F F 2.84 F
;
proc logistic data=alligator; class food gender;
model food=length / link=glogit;
output out=out p=p;
proc sort data=out; by _level_ length;
proc sgplot data=out;
title "Predicted probabilities";
series x=length y=p / group=_level_;
yaxis min=0 max=1;

```

5. Consider the data of Problem 8.10. Fit a proportional odds (cumulative logit) model with treatment and gender as predictors. Carefully interpret the treatment and gender effects in terms of odds ratios; include interpretation of 95% CIs.

```

data lung;
input therapy$ gender$ response count @@;
datalines;
s m 1 28 s m 2 45 s m 3 29 s m 4 26
s f 1 4 s f 2 12 s f 3 5 s f 4 2
a m 1 41 a m 2 44 a m 3 20 a m 4 20
a f 1 12 a f 2 7 a f 3 3 a f 4 1
;
proc logistic data=lung;
class therapy gender;
weight count;
model response=therapy gender;

```