

**HW #4 Solutions**

- 1 Problem 6.1 of Agresti. It results from high collinearity between weight and width variables.  $\text{logit}(\pi) = -9.35 + .834(\text{weight}) + .307(\text{width})$ .
- Like. ratio stat. = 32.9 (df = 2),  $P < .0001$ . There is extremely strong evidence that at least one variable affects the response.
  - Wald statistics are  $(.834/.671)^2 = 1.55$  and  $(.307=.182)^2 = 2.85$ . These each have df = 1, and the P-values are .21 and .09. These predictors are highly correlated (Pearson corr. = .887), so this is the problem of multicollinearity.

```
data crab;
infile 'crab2.dat' firstobs=2; * skip the first header line;
input color spine width satell weight;
  weight=weight/1000; color=color-1;
  if satell>0 then y=1; if satell=0 then y=0;
  if color=4 then light=0; if color < 4 then light=1;
run;

proc logistic data=crab descending;
model y=width weight ;
run;
* gives LR test stat = 32.9 for 2 df for H0: beta=0 ;

proc genmod data=crab descending; class color;
  model y = weight width / dist=bin link=logit lrci;
run;
* 95% LR CI includes 0 for both weight and width variable
```

- 2 Problem 6.2 of Agresti.

Run the following codes. Note the use of DO loops to construct combinations of factors

No (additional) effects met the 0.05 significance level for removal from the model.

All variables, religious attendance, political preference, attitude toward birth control, when positive, increase the odds of the person having a positive opinion toward birth\_control

```
data prob_6_2;
do political = 1 to 3;
do premarital = 1 to 2;
do religious = 1 to 2;
do birth_control = 1 to 2;
input count @;
output;
end;
end;
end;
datalines;
99 15 73 25 8 4 24 22
73 20 87 37 20 13 50 60
51 19 51 36 6 12 33 88
;
proc logistic descending; weight count;
model premarital = religious political birth_control /selection=backward fast;
```

```
run;
```

NOTE: No (additional) effects met the 0.05 significance level for removal from the model.

Analysis of Maximum Likelihood Estimates

Parameter	DF	Estimate	Standard Error	Wald Chi-Square	Pr > ChiSq
Intercept	1	-4.8739	0.3991	149.1532	<.0001
religious	1	1.1537	0.1695	46.3208	<.0001
political	1	0.3764	0.0990	14.4700	0.0001
birth_control	1	1.1345	0.1525	55.3242	<.0001

3 Problem 6.3 of Agresti. Run the following SAS codes.

Note that variables are coded as (1=Divorced, 2=Married), (1=Women, 2=Male), (1=yes, 2=No) etc.

Also note the design variables coded 1=Yes as 1 and 2=No as -1.

The odds of staying married decrease when a person says Yes to EMS or PMS. Also, the odds of staying married are lower for women compared to men.

At the same EMS level, positive attitude toward PMS increases the odds of staying married (strange)

```
data prob_6_3;
do married= 1 to 2;
  do gender = 1 to 2;
    do pms = 1 to 2;
      do ems = 1 to 2;
        input count @;
        output;
      end;
    end;
  end;
end;
datalines;
17 54 36 214 28 60 17 68
4 25 4 322 11 42 4 130
;
```

```
proc logistic descending; weight count; class ems pms gender;
model married = ems pms ems * pms gender;
run;
```

Class	Value	Design Variables
ems	1	1
	2	-1
pms	1	1
	2	-1
gender	1	1
	2	-1

Parameter	DF	Estimate	Standard Error	Wald Chi-Square	Pr > ChiSq
-----------	----	----------	----------------	-----------------	------------

Intercept		1	-0.7525	0.1283	34.4242	<.0001
ems	1	1	-0.7481	0.1285	33.8685	<.0001
pms	1	1	-0.0998	0.1302	0.5873	0.4435
ems*pms	1 1	1	0.4500	0.1282	12.3119	0.0005
gender	1	1	-0.1544	0.0729	4.4870	0.0342

## 4 Problem 6.5 of Agresti.

- a) The estimated odds of admission were 1.84 times higher for men than women. However, odds ratio for AG(D) = .90, so given department, the estimated odds of admission were .90 times as high for men as for women. Simpson's paradox strikes again! Men applied relatively more often to Departments A and B, whereas women applied relatively more often to Departments C, D, E, F. At the same time, admissions rates were relatively high for Departments A and B and relatively low for C, D, E, F. These two effects combine to give a relative advantage to men for admissions when we study the marginal association.
- b)
- c) The values of G2 are 2.68 for the model with no G effect and 2.56 for the model with G and D main effects. For the latter model, CI for conditional AG odds ratio is (0.87, 1.22).

## 5 Problem 6.8. Run the following SAS code.

Coded as: 'white-black', 'yes-no'

See the part of the output. There's strong dependency between victim's race and death penalty. If the victim's race is white, the odds of death penalty increase a lot (  $\exp(93)$  ) compared to when the victim's race is black.

```
data prob_6_8;
do victim = 1 to 2;
  do defendant= 1 to 2;
    do death = 1 to 2;
      input count @;
      output;
    end;
  end;
end;
datalines;
53 414 11 37 0 16 4 139
;
proc freq data=prob_6_8;
weight count;
table death * victim * defendant / cmh;
run;
```

## Cochran-Mantel-Haenszel Statistics (Based on Table Scores)

Statistic	Alternative Hypothesis	DF	Value	Prob
1	Nonzero Correlation	1	385.3371	<.0001
2	Row Mean Scores Differ	1	385.3371	<.0001
3	General Association	1	385.3371	<.0001

Estimates of the Common Relative Risk (Row1/Row2)

Type of Study	Method	Value	95% Confidence Limits	
Case-Control (Odds Ratio)	Mantel-Haenszel Logit **	100.3975 93.9186	54.0146 51.3184	186.6097 171.8819
Cohort (Col1 Risk)	Mantel-Haenszel Logit **	9.1546 8.8720	5.7358 5.6133	14.6109 14.0222
Cohort (Col2 Risk)	Mantel-Haenszel Logit	0.0943 0.1073	0.0699 0.0819	0.1272 0.1407

\*\* These logit estimators use a correction of 0.5 in every cell of those tables that contain a zero.