Solution to HW6

Problem 4.15

(a) The SAS program and part of the output are:

```
data prob4_15;
  input district $ n1-n4;
  datalines;
  NC 24 9 47 12
  NE 10 3 45 8
  NW 5 4 57 9
  SE 16 7 54 10
  SW 7 4 59 12
  ;

data prob4_15; set  prob4_15;
  array temp {4} n1-n4;

  do i=1 to 4;
    race = (i>2);                * race=0/1 for black/white;
    meritpay = (mod(i, 2) ne 0); * merit pay is 1 when i=1, 3;
    count=temp(i);
    output;
  end;
run;

proc freq data=prob4_15 order=data;
  weight count;
  tables district*race*meritpay / cmh;
run;
```

```
**********************************************************************

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

        Statistic    Alternative Hypothesis    DF       Value      Prob
        ---------------------------------------------------------------
            1        Nonzero Correlation        1      7.8149    0.0052
            2        Row Mean Scores Differ     1      7.8149    0.0052
            3        General Association        1      7.8149    0.0052
```

The CMH test of conditional independence between race and merit pay conditional on district is $\chi^2_{CMH} = 7.8$ with $df = 1$ and P-value $= 0.0052$, which indicates strong evidence against this conditional independence hypothesis.

Interpretation: In each district, the probabilities that white agents and black agents receive a merit pay are different.

(b) We can consider the following logistic regression model for the merit pay probability with main effects of race and district in the model:

$$\text{logit}\{P(\text{Merit Pay})\} = \alpha + \beta_1 \text{race} + \beta_2 D_{NC} + \beta_3 D_{NE} + \beta_4 D_{NW} + \beta_5 D_{SE},$$

where D's are dummy variables for districts. We can test the conditional independence hypothesis in (a) by testing $H_0 : \beta_1 = 0$. The SAS program and part of output are:

```
proc logistic;
  class district / param=ref;
```

```
  freq count;
  model meritpay (event="1") = race district;
run;

*************************************************************************

                    Analysis of Maximum Likelihood Estimates

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

    Intercept         1      0.7539             0.3639        4.2918        0.0383
    race              1      0.7913             0.2853        7.6914        0.0055
    district  NC      1    -0.00445             0.3849        0.0001        0.9908
    district  NE      1      0.2539             0.4367        0.3381        0.5609
    district  NW      1      0.1339             0.4162        0.1035        0.7476
    district  SE      1      0.1164             0.3950        0.0869        0.7682
```

The parameter estimate for β_1 is $\widehat{\beta}_1 = 0.7913$ with SE $= 0.2853$. The Wald test for $H_0 : \beta_1 = 0$ is $\chi^2 = 7.6914$ with $df = 1$ and P-value $= 0.0055$. Therefore, we reject the conditional independence hypothesis in (a) at significance level 0.05 (or any level > 0.0055). The conclusion is the same as that from the CMH test.

(c) Use the logistic model, we can get much more information than the CMH test. From the model, we can get the parameter estimate of the race effect to see which race is more likely to receive a merit pay. We can also see if there is a probability difference across different districts. Of course, we have to assume the underlying logistic model is correct. On the other hand, the CMH test does not assume the correct specification of the logistic model.

Problem 4.20

(a) A logistic regression model for Success probability with main effects of treatment and center can be implemented using the following SAS program (with part of the output):

```
data prob4_20;
  input center trt $ y y0;
  drug=(trt="Drug"); n = y+y0;
  datalines;
  1 Drug       11 25
  1 Control    10 27
  2 Drug       16 4
  2 Control    22 10
  3 Drug       14 5
  3 Control    7 12
  4 Drug       2 14
  4 Control    1 16
  5 Drug       6 11
  5 Control    0 12
  6 Drug       1 10
  6 Control    0 10
  7 Drug       1 4
  7 Control    1 8
  8 Drug       4 2
  8 Control    6 1
  ;

proc logistic;
  class center / param=ref;
  model y/n = drug center;
```

```
run;
```

```
*************************************************************************
```

Analysis of Maximum Likelihood Estimates

Parameter		DF	Estimate	Standard Error	Wald Chi-Square	Pr > ChiSq
Intercept		1	0.8859	0.6755	1.7201	0.1897
drug		1	0.7769	0.3067	6.4174	0.0113
center	1	1	-2.2079	0.7195	9.4166	0.0022
center	2	1	-0.1525	0.7381	0.0427	0.8363
center	3	1	-1.0550	0.7457	2.0015	0.1571
center	4	1	-3.6264	0.9071	15.9813	<.0001
center	5	1	-2.7278	0.8184	11.1104	0.0009
center	6	1	-4.3548	1.2293	12.5499	0.0004
center	7	1	-3.0056	1.0200	8.6836	0.0032

The output indicates that the treatment effect is significant at level 0.05 (P-value=0.0113 from a Wald test). The conditional odds of Success between the drug and the placebo conditional on center is estimated as $e^{0.7769} = 2.17$.

(b) The analysis of the treatment effect using a logistic regression model in (a) may not be valid due to many sparse cells in the data. Thus we used CMH to test the conditional independence of the treatment and response conditional on the center. The SAS program and relevant output are given in the following:

```
data prob4_20; set prob4_20;
  do i=1 to 2;
    success=2-i;
    if success=1 then
      count=y;
    else
      count=y0;
    output;
  end;
run;

proc freq data=prob4_20 order=data;
  weight count;
  tables center*trt*success / cmh;
run;
```

```
*************************************************************************
```

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

Statistic	Alternative Hypothesis	DF	Value	Prob
1	Nonzero Correlation	1	6.3841	0.0115
2	Row Mean Scores Differ	1	6.3841	0.0115
3	General Association	1	6.3841	0.0115

The CMH gives $\chi^2 = 6.3841$ with $df = 1$ so that P-value = 0.0115. Therefore, we reject the null hypothesis that the treatment and being success are conditionally independent give the center (at any level > 0.0115).

Problem 4.23

(a) The fitted model is

$$\text{logit}\{\widehat{P}(\text{cancer}|A,S,R)\} = -7 + 0.1A + 1.2S + 0.3R + 0.2R \times S.$$

Therefore, the logit of the cancer probability for blacks ($R = 1$) is

$$\text{logit}\{\widehat{P}(\text{cancer}|A,S,R=1)\} = -6.7 + 0.1A + 1.4S.$$

So the conditional odds-ratio given alcohol consumption between the response Y and S for blacks is $e^{1.4} = 4.06$.

The logit of the cancer probability for whites ($R = 0$) is

$$\text{logit}\{\widehat{P}(\text{cancer}|A,S,R=0)\} = -7 + 0.1A + 1.2S.$$

So the conditional odds-ratio given alcohol consumption between the response Y and S for whites is $e^{1.2} = 3.32$.

(b) We can interpret the coefficient of S using the logit model for whites:

$$\text{logit}\{\widehat{P}(\text{cancer}|A,S,R=0)\} = -7 + 0.1A + 1.2S.$$

It tells us the conditional effect of smoking for whites given alcohol consumption as we described at the end of (a). The P-value tests the conditional smoking effect (given alcohol consumption) for whites.

In order to interpret the coefficient of R, we can set $S = 0$ in the fitted equation to get

$$\text{logit}\{\widehat{P}(\text{cancer}|A,S=0,R)\} = -7 + 0.1A + 0.3R.$$

Therefore the coefficient 0.3 represents the race difference for non-smokers conditional on alcohol consumption. That is, the conditional odds ratio of cancer between non-smoking blacks and non-smoking whites given the alcohol consumption is $e^{0.3} = 1.35$. The P-value tests the conditional race effect (given alcohol consumption) for non-smokers.

(c) Now the new prediction equation is

$$\text{logit}\{\widehat{P}(\text{cancer}|A,S,R)\} = \widehat{\alpha} + \widehat{\beta}_1 A + \widehat{\beta}_2 S + \widehat{\beta}_3 R + \widehat{\beta}_4 R \times S + 0.04A \times R.$$

The predictioin equation for blacks is

$$\text{logit}\{\widehat{P}(\text{cancer}|A,S,R=1)\} = \widehat{\alpha}+\widehat{\beta}_1 A+\widehat{\beta}_2 S+\widehat{\beta}_3+\widehat{\beta}_4 S+0.04A = (\widehat{\alpha}+\widehat{\beta}_3)+(\widehat{\beta}_1+0.04)A+(\widehat{\beta}_2+\widehat{\beta}_4)S.$$

So the alcohol effect for blacks is $\widehat{\beta}_1 + 0.04$.

The predictioin equation for whites is

$$\text{logit}\{\widehat{P}(\text{cancer}|A, S, R = 0)\} = \widehat{\alpha} + \widehat{\beta}_1 A + \widehat{\beta}_2 S$$

So the alcohol effect for whites is $\widehat{\beta}_1$, and 0.04 is the difference between these two alcohol effects for blacks and whites.

Problem 4.24

(a) The SAS program and relevant output for the main effect model are:

```
data prob4_24;
  input PTID D T Y @@;
  datalines;
  1 45 0 0 13 50 1 0 25 20 1 0
  2 15 0 0 14 75 1 1 26 45 0 1
  3 40 0 1 15 30 0 0 27 15 1 0
  4 83 1 1 16 25 0 1 28 25 0 1
  5 90 1 1 17 20 1 0 29 15 1 0
  6 25 1 1 18 60 1 1 30 30 0 1
  7 35 0 1 19 70 1 1 31 40 0 1
  8 65 0 1 20 30 0 1 32 15 1 0
  9 95 0 1 21 60 0 1 33 135 1 1
  10 35 0 1 22 61 0 0 34 20 1 0
  11 75 0 1 23 65 0 1 35 40 1 0
  12 45 1 1 24 15 1 0
;

proc logistic descending;
  model y = d t;
run;
```

```
*************************************************************************
                    Analysis of Maximum Likelihood Estimates

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

      Intercept     1     -1.4173         1.0946        1.6766        0.1954
      D             1      0.0687         0.0264        6.7611        0.0093
      T             1     -1.6589         0.9229        3.2314        0.0722
```

From the output, we see that the duration of the surgery has a significant effect give type of the device (P-value=0.0093). Given the same type of device, with 10 min increase of the duration of the surgery, the odds of experiencing a sore throat almost double ($e^{0.687} \approx 2$).

The device has a marginally significant effect given the duration of the surgery (P-value=0.0722). For patients with the same amount of surgery time, the odds of experiencing a sore throat for those treated with tracheal tube is only 0.19 times the odds of experiencing a sore throat for those treated with laryngeal mask.

(b) The Wald test for duration effect produces $\chi^2 = 6.7611$ with $df = 1$ and P-value = 0.0093. Significant at level 0.05.

(c) The logistic model with interaction of duration and type of device is:

```
proc genmod descending;
  model y = d t d*t / dist=bin type3;
run;
```

```
*******************************************************************************

              Analysis Of Maximum Likelihood Parameter Estimates

                          Standard       Wald 95%            Wald
 Parameter  DF  Estimate     Error   Confidence Limits  Chi-Square  Pr > ChiSq

 Intercept   1    0.0498    1.4694   -2.8302    2.9298        0.00      0.9730
 D           1    0.0285    0.0343   -0.0387    0.0957        0.69      0.4062
 T           1   -4.4722    2.4671   -9.3076    0.3631        3.29      0.0699
 D*T         1    0.0746    0.0578   -0.0386    0.1878        1.67      0.1966
 Scale       0    1.0000    0.0000    1.0000    1.0000

NOTE: The scale parameter was held fixed.

                     LR Statistics For Type 3 Analysis

                                          Chi-
                Source           DF     Square    Pr > ChiSq

                D                 1       0.81        0.3690
                T                 1       3.95        0.0470
                D*T               1       1.82        0.1777
```

The prediction equation is

$$\text{logit}\{P(\text{sore throat}|D,T)\} = 0.0498 + 0.0285D - 4.4722T + 0.0746D \times T.$$

The prediction equation for $T = 1$ is:

$$\text{logit}\{P(\text{sore throat}|D,T=1)\} = 0.0498+0.0285D-4.4722+0.0746D = -4.4224+0.1031D.$$

That is, for patients using tracheal tube, with 10 min increase of surgery time, the odds of experiencing a sore throat multiplies by 2.8.

The prediction equation for $T = 0$ is:

$$\text{logit}\{P(\text{sore throat}|D,T=0)\} = 0.0498 + 0.0285D.$$

That is, for patients using laryngeal mask, with 10 min increase of surgery time, the odds of experiencing a sore throat multiplies by 1.33.

(d) The Wald test for the interaction $D \times T$: $\chi^2 = 1.67$ with $df = 1$ so P-value=0.1966, not significant!

The LRT for the interaction $D \times T$: $G^2 = 1.82$ with $df = 1$, so the P-value=0.1777, not significant either. From both the Wald test and the LRT, it seems that we can remove the interaction term $D \times T$ from the model.