STAT436 **Homework 8** **Due Tuesday 02/25 at 2.30pm**

Directions: Showing all work, answer the following three exercises below. As always, conclusions and justifications are to be given in clear detailed English. Please type up your solutions or write very neatly.

1. [***PO model***] Samuels & Witmer (*Statistics for the Life Sciences*,1999:429) present an example in which 50 patients were randomized to receive either of pain medication A or B (25 patients in each group), and the measured dependent variable was the response to the pain medication in terms of pain relief. For the levels of Y, the researchers simply recorded as 1 (for "None"), 2 (for "Some"), 3 (for "Substantial"), or 4 (for "Complete). The counts are in this table below, and our goal here is to decide if the drugs differ in terms of pain relief (and if so, how).

	PAIN RELIEF				
	None (Y = 1)	**Some (Y = 2)**	**Substantial (Y = 3)**	**Complete (Y = 4)**	**Total**
Drug A	**3**	**7**	**10**	**5**	**25**
Drug B	**7**	**11**	**5**	**2**	**25**

   (a) Examine the results of the "**`proc freq`**" and discuss your findings, bearing in mind all necessary assumptions of the tests. For example, the MH test shows significance (p = 0.0282), but the Chi-square (p = 0.1422) and FET (p = 0.1618) tests show no significance – comment on the relevance of these tests and result here. Hint: you may want to search the Web or a basic statistics textbook to discover why **neither** the usual Chi-square nor the FET test is appropriate for these data, and why the MH Chi-square test **is appropriate**.
   (b) Examine the results from the **`proc logistic`** and discuss your findings. ***Remember to give all assumptions here!*** Do all necessary assumptions seem to be met here (check the proportional odds assumption)?
   (c) Based on this (**`proc logistic`**) output, do you feel the drugs differ in terms of pain relief?
   (d) Write down the predictive formulas here.
   (e) Clearly interpret the odds ratio for this **`proc logistic`** fit.
   (f) Use the predictive formulas in part (d) to give the predicted values in each one of the cells in the table above and compare the PO predicted values with the actual values and the expected values using the **`proc freq`** chi-square method.

2. [***Logistic model – two groups***] On p. 113, Collett (*Modelling Binary Data*, Chapman & Hall, 2nd edition, 2003) describes an insecticide toxicity study in which flour beetles, *Tribolium castaneum*, were sprayed with one of three different insecticides in solution in Shell oil P31. The three insecticides used were dichloro-diphenyltrichloroethane (DDT) at 2.0% w/v, γ-benzene hexachloride (γ-BHC) used at 1.5% w/v, and a mixture of the two. In the experiment, batches of about fifty insects were exposed to varying deposits of spray, measured in units of mg/10 cm^2. The resulting data on the proportion of insects killed after a period of six days are given in the Table below. In modelling these data, the (natural) logarithm of the amount of deposit of insecticide is used as the explanatory variable in a linear logistic model, and the deposit levels were 2.00, 2.64, 3.48, 4.59, 6.06, and 8.00 (mg/10 cm^2).

	deposit levels					
Insecticide	**2.00**	**2.64**	**3.48**	**4.59**	**6.06**	**8.00**
DDT	**3 of 50**	**5 of 49**	**19 of 47**	**19 of 38**	**24 of 49**	**35 of 50**
g-BHC	**2 of 50**	**14 of 49**	**20 of 50**	**27 of 50**	**41 of 50**	**40 of 50**
mixture	**28 of 50**	**37 of 50**	**46 of 50**	**48 of 50**	**48 of 50**	**50 of 50**

Our goal here is to examine and compare only the '**DDT**' and the '**mixture**' (combination of both 'DDT and γ-BHC') treatments using the results given in the SAS program and output given below.

(a) Write down the model function fit in Output 2B and **give all necessary assumptions for the model**.
(b) Using Output 2B, write down the predictive logistic formula for each of the two groups – one predicted formula for the '**DDT**' group and one for the '**mixture**' group.
(c) Using Output 2B, give your estimates of the $\mathbf{LD_{50}}$'s (***remember to give these on the original scale!***) for each of the two groups.
(d) Using Outputs 2B and 2C, test whether the two treatments share a common slope parameter (with different intercepts). Use the **BEST** test – meaning, do not use the 'Wald Chi-square' test statistic here. Showing your calculations, report the hypotheses, calculated test statistic, p-value, and clear conclusion.
(e) Assuming common slopes (i.e., using Output 2C), test whether the intercepts differ (the Wald test will have to suffice here since we did not run the new "reduced" model). Again, report the hypotheses, calculated test statistic, p-value, and clear conclusion.
(f) Briefly comment on the quality of the fit of the model in Output 2C.

3. [***Logistic model – three groups***] Repeat each of the steps in Exercise 2 but using the full dataset for all three treatment groups – the data are analyzed in Outputs 3A and 3B below.

Homework 8 Appendix

SAS Program and Output 1

```
data one;
  do resp='1_none','2_some',
     '3_subs','4_comp';
     resp2=1*(resp='1_none')
          +2*(resp='2_some')
          +3*(resp='3-subs')
          +4*(resp='4-comp');
  do drug='A','B';
    dummy_drug_a=(drug='A');
    input count @@; output;
  end; end; cards;
3 7 7 11 10 5 5 2
;
```

RESP	RESP2	DRUG	DRUGA	count
1_none	1	A	1	3
1_none	1	B	0	7
2_some	2	A	1	7
2_some	2	B	0	11
3_subs	3	A	1	10
3_subs	3	B	0	5
4_comp	4	A	1	5
4_comp	4	B	0	2

The FREQ Procedure
Table of drug by RESP

```
proc print noobs;
run;
proc freq;
  weight count;
  tables drug*resp/
    chisq fisher
    nopercent norow expected;
run;
```

DRUG RESP

Frequency / Expected / Col Pct	1_none	2_some	3_subs	4_comp	Total
A	3	7	10	5	25
	5	9	7.5	3.5	
	30.00	38.89	66.67	71.43	
B	7	11	5	2	25
	5	9	7.5	3.5	
	70.00	61.11	33.33	28.57	
Total	10	18	15	7	50

Statistics for Table of DRUG by RESP

Statistic	DF	Value	Prob
Chi-Square	3	5.4413	0.1422
Likelihood Ratio Chi-Square	3	5.5693	0.1346
Mantel-Haenszel Chi-Square	1	4.8165	0.0282
Phi Coefficient		0.3299	
Contingency Coefficient		0.3133	
Cramer's V		0.3299	

WARNING: 25% of the cells have expected counts less than 5. Chi-Square may not be a valid test.

Fisher's Exact Test

Table Probability (P)	0.0019
Pr <= P	0.1618

Sample Size = 50

SAS Program/Output 1 continued

```
proc logistic;
  weight count;
  model RESP2=
    dummy_drug_a;
run;
```

The LOGISTIC Procedure

Model Information

Data Set	WORK.ONE
Response Variable	RESP2
Number of Response Levels	4
Weight Variable	count
Model	cumulative logit
Optimization Technique	Fisher's scoring
Number of Observations Read	8
Number of Observations Used	8
Sum of Weights Read	50
Sum of Weights Used	50

Response Profile

Ordered Value	RESP2	Total Frequency	Total Weight

```
         1            1            2        10.000000
         2            2            2        18.000000
         3            3            2        15.000000
         4            4            2         7.000000
Probabilities modeled are cumulated over the lower Ordered Values.

    Score Test for the Proportional Odds Assumption
         Chi-Square       DF     Pr > ChiSq
             0.2683        2         0.8745

              Model Fit Statistics
                   Intercept    Intercept and
    Criterion          Only       Covariates
    AIC             138.613          135.302
    SC              138.851          135.620
    -2 Log L        132.613          127.302

         Testing Global Null Hypothesis: BETA=0
Test                 Chi-Square       DF     Pr > ChiSq
Likelihood Ratio         5.3107        1         0.0212
Score                    5.1209        1         0.0236
Wald                     5.0910        1         0.0241

             Analysis of Maximum Likelihood Estimates
                                Standard          Wald
Parameter     DF   Estimate      Error    Chi-Square    Pr > ChiSq
Intercept 1    1    -0.8957     0.4135        4.6924        0.0303
Intercept 2    1     0.8718     0.4114        4.4914        0.0341
Intercept 3    1     2.5625     0.5445       22.1448        <.0001
dummy_drug_a   1    -1.2205     0.5409        5.0910        0.0241

                     Odds Ratio Estimates
                           Point          95% Wald
        Effect          Estimate      Confidence Limits
        dummy_drug_a       0.295      0.102       0.852
```

SAS Program and Output 2A

```
data insect;
  do type='a DDT     ','b mixture';
    dummy_mixture=(type='b mixture');
  do deposit=2,2.64,3.48,4.59,6.06,8;
    log_deposit=log(deposit);
    input dead n @@; output;
  end; end; datalines;
3 50 5 49 19 47 19 38 24 49 35 50
28 50 37 50 46 50 48 50 48 50 50 50
;
proc print noobs;
  var type log_deposit dead n
  dummy_mixture;
run;
```

type	log_deposit	dead	n	dummy_mixture
a DDT	0.69315	3	50	0
a DDT	0.97078	5	49	0
a DDT	1.24703	19	47	0
a DDT	1.52388	19	38	0
a DDT	1.80171	24	49	0
a DDT	2.07944	35	50	0
b mixture	0.69315	28	50	1
b mixture	0.97078	37	50	1
b mixture	1.24703	46	50	1
b mixture	1.52388	48	50	1
b mixture	1.80171	48	50	1
b mixture	2.07944	50	50	1

SAS Program 2B

```
proc logistic;
  model dead/n=log_deposit dummy_mixture dummy_mixture*log_deposit;
run;
```

SAS Output 2B

The LOGISTIC Procedure

Model Information

Data Set	WORK.INSECT
Response Variable (Events)	DEAD
Response Variable (Trials)	n
Model	binary logit
Optimization Technique	Fisher's scoring
Number of Observations Read	12
Number of Observations Used	12
Sum of Frequencies Read	583
Sum of Frequencies Used	583

Response Profile

Ordered Value	Binary Outcome	Total Frequency
1	Event	362
2	Nonevent	221

Model Fit Statistics

Criterion	Intercept Only	Intercept and Covariates
AIC	775.768	506.586
SC	780.136	524.059
-2 Log L	773.768	498.586

Testing Global Null Hypothesis: BETA=0

Test	Chi-Square	DF	Pr > ChiSq
Likelihood Ratio	275.1817	3	<.0001
Score	230.9021	3	<.0001
Wald	133.8380	3	<.0001

Analysis of Maximum Likelihood Estimates

Parameter	DF	Estimate	Standard Error	Wald Chi-Square	Pr > ChiSq
Intercept	1	-3.8308	0.5002	58.6591	<.0001
log_deposit	1	2.2824	0.3190	51.2033	<.0001
dummy_mixture	1	1.7102	0.7741	4.8814	0.0271
log_depos*dummy_mixt	1	1.1058	0.6561	2.8407	0.0919

SAS Program 2C

```
proc logistic;
  model dead/n=log_deposit dummy_mixture / lackfit;
run;
```

SAS Output 2C

The LOGISTIC Procedure

Model Information

Data Set	WORK.INSECT
Response Variable (Events)	DEAD
Response Variable (Trials)	n
Model	binary logit
Optimization Technique	Fisher's scoring

Number of Observations Read	12
Number of Observations Used	12
Sum of Frequencies Read	583
Sum of Frequencies Used	583

Response Profile

Ordered Value	Binary Outcome	Total Frequency
1	Event	362
2	Nonevent	221

Model Fit Statistics

Criterion	Intercept Only	Intercept and Covariates
AIC	775.768	507.663
SC	780.136	520.768
-2 Log L	773.768	501.663

Testing Global Null Hypothesis: BETA=0

Test	Chi-Square	DF	Pr > ChiSq
Likelihood Ratio	272.1047	2	<.0001
Score	227.2781	2	<.0001
Wald	146.2932	2	<.0001

Analysis of Maximum Likelihood Estimates

Parameter	DF	Estimate	Standard Error	Wald Chi-Square	Pr > ChiSq
Intercept	1	-4.3002	0.4446	93.5706	<.0001
log_deposit	1	2.5927	0.2770	87.6127	<.0001
dummy_mixture	1	2.9823	0.2633	128.3194	<.0001

Odds Ratio Estimates

Effect	Point Estimate	95% Wald Confidence Limits	
log_deposit	13.365	7.766	23.002
dummy_mixture	19.734	11.779	33.061

Hosmer and Lemeshow Goodness-of-Fit Test

Chi-Square	DF	Pr > ChiSq
13.6718	10	0.1885

SAS Program 3

```
data insect;
  do type='a DDT    ','b g-BHC  ','c mixture';
    dummy_ddt=(type='a DDT    ');
    dummy_mixture=(type='c mixture');
  do deposit=2,2.64,3.48,4.59,6.06,8;
    log_deposit=log(deposit);
    input dead n @@; output;
  end; end; datalines;
3 50 5 49 19 47 19 38 24 49 35 50
2 50 14 49 20 50 27 50 41 50 40 50
28 50 37 50 46 50 48 50 48 50 50 50
;
proc logistic;
  model dead/n=log_deposit dummy_ddt dummy_mixture
     dummy_ddt*log_deposit dummy_mixture*log_deposit;
run;
proc logistic;
  model dead/n= log_deposit dummy_ddt
     dummy_mixture/lackfit;
run;
```

SAS Output 3 – First Logistic

```
                    The LOGISTIC Procedure

                                Model Information
        Data Set                          WORK.INSECT
        Response Variable (Events)        dead
        Response Variable (Trials)        n
        Model                             binary logit
        Optimization Technique            Fisher's scoring

        Number of Observations Read          18
        Number of Observations Used          18
        Sum of Frequencies Read             882
        Sum of Frequencies Used             882

                     Response Profile
              Ordered     Binary          Total
                Value     Outcome     Frequency
                    1     Event             506
                    2     Nonevent          376

                    Model Fit Statistics
                                          Intercept
                               Intercept        and
              Criterion             Only Covariates
              AIC                1205.481    827.644
              SC                 1210.263    856.337
              -2 Log L           1203.481    815.644
```

Testing Global Null Hypothesis: BETA=0

Test	Chi-Square	DF	Pr > ChiSq
Likelihood Ratio	387.8364	5	<.0001
Score	327.4879	5	<.0001
Wald	205.5601	5	<.0001

Analysis of Maximum Likelihood Estimates

Parameter	DF	Estimate	Standard Error	Wald Chi-Square	Pr > ChiSq
Intercept	1	-4.0428	0.4972	66.1070	<.0001
log_deposit	1	2.8381	0.3392	69.9995	<.0001
dummy_ddt	1	0.2120	0.7053	0.0904	0.7637
dummy_mixture	1	1.9223	0.7722	6.1977	0.0128
log_deposi*dummy_ddt	1	-0.5557	0.4656	1.4241	0.2327
log_depos*dummy_mixt	1	0.5500	0.6661	0.6818	0.4090

SAS Output 3 – Second Logistic

The LOGISTIC Procedure

Model Information

Data Set	WORK.INSECT
Response Variable (Events)	dead
Response Variable (Trials)	n
Model	binary logit
Optimization Technique	Fisher's scoring

Number of Observations Read	18
Number of Observations Used	18
Sum of Frequencies Read	882
Sum of Frequencies Used	882

Response Profile

Ordered Value	Binary Outcome	Total Frequency
1	Event	506
2	Nonevent	376

Model Fit Statistics

Criterion	Intercept Only	Intercept and Covariates
AIC	1205.481	827.037
SC	1210.263	846.166
-2 Log L	1203.481	819.037

Testing Global Null Hypothesis: BETA=0

Test	Chi-Square	DF	Pr > ChiSq
Likelihood Ratio	384.4439	3	<.0001
Score	317.1755	3	<.0001
Wald	217.7538	3	<.0001

Analysis of Maximum Likelihood Estimates

Parameter	DF	Estimate	Standard Error	Wald Chi-Square	Pr > ChiSq
Intercept	1	-3.8396	0.3313	134.3314	<.0001
log_deposit	1	2.6937	0.2146	157.5363	<.0001
dummy_ddt	1	-0.6144	0.1999	9.4516	0.0021
dummy_mixture	1	2.4169	0.2379	103.1901	<.0001

Odds Ratio Estimates

Effect	Point Estimate	95% Wald Confidence Limits	
log_deposit	14.786	9.709	22.519
dummy_ddt	0.541	0.366	0.800
dummy_mixture	11.211	7.033	17.872

Hosmer and Lemeshow Goodness-of-Fit Test

Chi-Square	DF	Pr > ChiSq
11.0284	7	0.1374