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).

<table>
<thead>
<tr>
<th>PAIN RELIEF</th>
<th>None (Y = 1)</th>
<th>Some (Y = 2)</th>
<th>Substantial (Y = 3)</th>
<th>Complete (Y = 4)</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Drug A</td>
<td>3</td>
<td>7</td>
<td>10</td>
<td>5</td>
<td>25</td>
</tr>
<tr>
<td>Drug B</td>
<td>7</td>
<td>11</td>
<td>5</td>
<td>2</td>
<td>25</td>
</tr>
</tbody>
</table>

(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 cm². 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 cm²).
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 LD50’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.

<table>
<thead>
<tr>
<th>Insecticide</th>
<th>2.00</th>
<th>2.64</th>
<th>3.48</th>
<th>4.59</th>
<th>6.06</th>
<th>8.00</th>
</tr>
</thead>
<tbody>
<tr>
<td>DDT</td>
<td>3 of 50</td>
<td>5 of 49</td>
<td>19 of 47</td>
<td>19 of 38</td>
<td>24 of 49</td>
<td>35 of 50</td>
</tr>
<tr>
<td>γ-BHC</td>
<td>2 of 50</td>
<td>14 of 49</td>
<td>20 of 50</td>
<td>27 of 50</td>
<td>41 of 50</td>
<td>40 of 50</td>
</tr>
<tr>
<td>mixture</td>
<td>28 of 50</td>
<td>37 of 50</td>
<td>46 of 50</td>
<td>48 of 50</td>
<td>48 of 50</td>
<td>50 of 50</td>
</tr>
</tbody>
</table>

---

**Homework 8 Appendix**

**SAS Program and Output 1**

```sas
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
```
### SAS Code/Output 1 continued

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

### Frequency Table

<table>
<thead>
<tr>
<th>DRUG</th>
<th>RESP</th>
<th>1_none</th>
<th>2_some</th>
<th>3_subs</th>
<th>4_comp</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>25</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>7</td>
<td>10</td>
<td>5</td>
<td>30.00</td>
<td>38.89</td>
</tr>
<tr>
<td>B</td>
<td>7</td>
<td>11</td>
<td>5</td>
<td>2</td>
<td>5</td>
<td>9</td>
</tr>
<tr>
<td></td>
<td>5</td>
<td>9</td>
<td>7.5</td>
<td>3.5</td>
<td>70.00</td>
<td>61.11</td>
</tr>
<tr>
<td>Total</td>
<td>10</td>
<td>18</td>
<td>15</td>
<td>7</td>
<td>50</td>
<td></td>
</tr>
</tbody>
</table>

### Statistics for Table of DRUG by RESP

<table>
<thead>
<tr>
<th>Statistic</th>
<th>DF</th>
<th>Value</th>
<th>Prob</th>
</tr>
</thead>
<tbody>
<tr>
<td>Chi-Square</td>
<td>3</td>
<td>5.4413</td>
<td>0.1422</td>
</tr>
<tr>
<td>Likelihood Ratio Chi-Square</td>
<td>3</td>
<td>5.5693</td>
<td>0.1346</td>
</tr>
<tr>
<td>Mantel-Haenszel Chi-Square</td>
<td>1</td>
<td>4.8165</td>
<td>0.0282</td>
</tr>
<tr>
<td>Phi Coefficient</td>
<td></td>
<td>0.3299</td>
<td></td>
</tr>
<tr>
<td>Contingency Coefficient</td>
<td></td>
<td>0.3133</td>
<td></td>
</tr>
<tr>
<td>Cramer's V</td>
<td></td>
<td>0.3299</td>
<td></td>
</tr>
</tbody>
</table>

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

Fisher's Exact Test

<table>
<thead>
<tr>
<th>Table Probability (P)</th>
<th>Pr &lt;= P</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.0019</td>
<td>0.1618</td>
</tr>
</tbody>
</table>

Sample Size = 50

### 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

<table>
<thead>
<tr>
<th>Ordered Value</th>
<th>Total RESP2</th>
<th>Total Frequency</th>
<th>Total Weight</th>
</tr>
</thead>
</table>
Proportions modeled are cumulated over the lower Ordered Values.

Score Test for the Proportional Odds Assumption

<table>
<thead>
<tr>
<th>Chi-Square</th>
<th>DF</th>
<th>Pr &gt; ChiSq</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.2683</td>
<td>2</td>
<td>0.8745</td>
</tr>
</tbody>
</table>

Model Fit Statistics

<table>
<thead>
<tr>
<th>Criterion</th>
<th>Intercept Only</th>
<th>Intercept and Covariates</th>
</tr>
</thead>
<tbody>
<tr>
<td>AIC</td>
<td>138.613</td>
<td>135.302</td>
</tr>
<tr>
<td>SC</td>
<td>138.851</td>
<td>135.620</td>
</tr>
<tr>
<td>-2 Log L</td>
<td>132.613</td>
<td>127.302</td>
</tr>
</tbody>
</table>

Testing Global Null Hypothesis: BETA=0

<table>
<thead>
<tr>
<th>Test</th>
<th>Chi-Square</th>
<th>DF</th>
<th>Pr &gt; ChiSq</th>
</tr>
</thead>
<tbody>
<tr>
<td>Likelihood Ratio</td>
<td>5.3107</td>
<td>1</td>
<td>0.0212</td>
</tr>
<tr>
<td>Score</td>
<td>5.1209</td>
<td>1</td>
<td>0.0236</td>
</tr>
<tr>
<td>Wald</td>
<td>5.0910</td>
<td>1</td>
<td>0.0241</td>
</tr>
</tbody>
</table>

Analysis of Maximum Likelihood Estimates

<table>
<thead>
<tr>
<th>Parameter</th>
<th>DF</th>
<th>Estimate</th>
<th>Error</th>
<th>Chi-Square</th>
<th>Pr &gt; ChiSq</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept 1</td>
<td>1</td>
<td>-0.8957</td>
<td>0.4135</td>
<td>4.6924</td>
<td>0.0303</td>
</tr>
<tr>
<td>Intercept 2</td>
<td>1</td>
<td>0.8718</td>
<td>0.4114</td>
<td>4.4914</td>
<td>0.0341</td>
</tr>
<tr>
<td>Intercept 3</td>
<td>1</td>
<td>2.5625</td>
<td>0.5445</td>
<td>22.1448</td>
<td>&lt;.0001</td>
</tr>
<tr>
<td>dummy_drug_a</td>
<td>1</td>
<td>-1.2205</td>
<td>0.5409</td>
<td>5.0910</td>
<td>0.0241</td>
</tr>
</tbody>
</table>

Odds Ratio Estimates

<table>
<thead>
<tr>
<th>Effect</th>
<th>Estimate</th>
<th>95% Wald Confidence Limits</th>
</tr>
</thead>
<tbody>
<tr>
<td>dummy_drug_a</td>
<td>0.295</td>
<td>0.102 0.852</td>
</tr>
</tbody>
</table>

SAS Program and Output 2A

```sas
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;
```

```sas
  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 48 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

```sas
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 Binary Total
Value Outcome Frequency
1 Event 362
2 Nonevent 221

Model Fit Statistics
 Criterion Only 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 Standard Wald
 DF Estimate Error 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_deposit*dummy_mixture 1 1.1058 0.6561 2.8407 0.0919
```

SAS Program 2C

```sas
proc logistic;
   model dead/n=log_deposit dummy_mixture / lackfit;
run;
```
The LOGISTIC Procedure

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 Binary Total
Value Outcome Frequency
1 Event 362
2 Nonevent 221

Model Fit Statistics

Intercept and Criterion
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  Error  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  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

```sas
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 40 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
```

Page 7 of 9
Testing Global Null Hypothesis: BETA=0

<table>
<thead>
<tr>
<th>Test</th>
<th>Chi-Square</th>
<th>DF</th>
<th>Pr &gt; ChiSq</th>
</tr>
</thead>
<tbody>
<tr>
<td>Likelihood Ratio</td>
<td>387.8364</td>
<td>5</td>
<td>&lt;.0001</td>
</tr>
<tr>
<td>Score</td>
<td>327.4879</td>
<td>5</td>
<td>&lt;.0001</td>
</tr>
<tr>
<td>Wald</td>
<td>205.5601</td>
<td>5</td>
<td>&lt;.0001</td>
</tr>
</tbody>
</table>

Analysis of Maximum Likelihood Estimates

<table>
<thead>
<tr>
<th>Parameter</th>
<th>DF</th>
<th>Standard Estimate</th>
<th>Wald Estimate</th>
<th>Chi-Square</th>
<th>Pr &gt; ChiSq</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept</td>
<td>1</td>
<td>-4.0428</td>
<td>0.4972</td>
<td>66.1070</td>
<td>&lt;.0001</td>
</tr>
<tr>
<td>log_deposit</td>
<td>1</td>
<td>2.8381</td>
<td>0.3392</td>
<td>69.9995</td>
<td>&lt;.0001</td>
</tr>
<tr>
<td>dummy_ddt</td>
<td>1</td>
<td>0.2120</td>
<td>0.7053</td>
<td>0.0904</td>
<td>0.7637</td>
</tr>
<tr>
<td>dummy_mixture</td>
<td>1</td>
<td>1.9223</td>
<td>0.7722</td>
<td>6.1977</td>
<td>0.0128</td>
</tr>
<tr>
<td>log_deposit*dummy_ddt</td>
<td>1</td>
<td>-0.5557</td>
<td>0.4656</td>
<td>1.4241</td>
<td>0.2327</td>
</tr>
<tr>
<td>log_deposit*dummy_mixt</td>
<td>1</td>
<td>0.5500</td>
<td>0.6661</td>
<td>0.8818</td>
<td>0.4090</td>
</tr>
</tbody>
</table>

SAS Output 3 – Second Logistic

The LOGISTIC Procedure

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

<table>
<thead>
<tr>
<th>Ordered Value</th>
<th>Binary Outcome</th>
<th>Total Frequency</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Event</td>
<td>506</td>
</tr>
<tr>
<td>2</td>
<td>Nonevent</td>
<td>376</td>
</tr>
</tbody>
</table>

Model Fit Statistics

<table>
<thead>
<tr>
<th>Criterion</th>
<th>Intercept and Only Covariates</th>
</tr>
</thead>
<tbody>
<tr>
<td>AIC</td>
<td>1205.481</td>
</tr>
<tr>
<td>SC</td>
<td>1210.263</td>
</tr>
<tr>
<td>-2 Log L</td>
<td>1203.481</td>
</tr>
</tbody>
</table>

Testing Global Null Hypothesis: BETA=0

<table>
<thead>
<tr>
<th>Test</th>
<th>Chi-Square</th>
<th>DF</th>
<th>Pr &gt; ChiSq</th>
</tr>
</thead>
<tbody>
<tr>
<td>Likelihood Ratio</td>
<td>384.4439</td>
<td>3</td>
<td>&lt;.0001</td>
</tr>
<tr>
<td>Score</td>
<td>317.1755</td>
<td>3</td>
<td>&lt;.0001</td>
</tr>
<tr>
<td>Wald</td>
<td>217.7538</td>
<td>3</td>
<td>&lt;.0001</td>
</tr>
</tbody>
</table>
### Analysis of Maximum Likelihood Estimates

<table>
<thead>
<tr>
<th>Parameter</th>
<th>DF</th>
<th>Estimate</th>
<th>Error</th>
<th>Chi-Square</th>
<th>Pr &gt; ChiSq</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept</td>
<td>1</td>
<td>-3.8396</td>
<td>0.3313</td>
<td>134.3314</td>
<td>&lt;.0001</td>
</tr>
<tr>
<td>log_deposit</td>
<td>1</td>
<td>2.6937</td>
<td>0.2146</td>
<td>157.5363</td>
<td>&lt;.0001</td>
</tr>
<tr>
<td>dummy_ddt</td>
<td>1</td>
<td>-0.6144</td>
<td>0.1999</td>
<td>9.4516</td>
<td>0.0021</td>
</tr>
<tr>
<td>dummy_mixture</td>
<td>1</td>
<td>2.4169</td>
<td>0.2379</td>
<td>103.1901</td>
<td>&lt;.0001</td>
</tr>
</tbody>
</table>

### Odds Ratio Estimates

<table>
<thead>
<tr>
<th>Effect</th>
<th>Point</th>
<th>95% Wald</th>
</tr>
</thead>
<tbody>
<tr>
<td>log_deposit</td>
<td>14.786</td>
<td>9.709</td>
</tr>
<tr>
<td>dummy_ddt</td>
<td>0.541</td>
<td>0.366</td>
</tr>
<tr>
<td>dummy_mixture</td>
<td>11.211</td>
<td>7.033</td>
</tr>
</tbody>
</table>

### Hosmer and Lemeshow Goodness-of-Fit Test

<table>
<thead>
<tr>
<th>Chi-Square</th>
<th>DF</th>
<th>Pr &gt; ChiSq</th>
</tr>
</thead>
<tbody>
<tr>
<td>11.0284</td>
<td>7</td>
<td>0.1374</td>
</tr>
</tbody>
</table>