Final Project

Britt Kravets

Due Date: 04/30/2023 by 11:59 PM EST

Please note that the full R code and output is included in the Appendix at the end of this report.

In this project, we used three data sets from a study of a two-level Road Departure Avoidance System at Horizontal Curves (RDAS-HC). The participants finished a horizontal curves driving test under various circumstances such as low visibility and wet pavement conditions. Researchers examined whether the RDAS-HC system can reduce the probability of lane departures in horizontal curves and whether the RDAS-HC system distracts drivers. The variables included were Distraction (from -3 to 3), Lane Departure (yes, no), Level of RDAS-SC (0,1,2), Male (gender: male, female), Night (illumination: daytime, night), and Wet (pavement condition: dry, wet).

Problem 1

In problem one I used Data1.dat. In problem one we wanted to use odds ratios to determine some information about how level and gender effected lane departure. I made a 2x2x2 table in order to visualize the data (**Table 1**). Next I calculated the conditional odds ratios giving the effectiveness of reducing lane departure occurred for males compared to females for each level of RDAS-HC system, and the marginal odds ratio between gender and lane departure (**Table 2**). For level 0, males were 1.9 times less likely to have lane deviation than females, for level 1 males were 2.6 times less likely to have lane deviation than females, and for level 2 males were 1.7 times less likely to have lane deviation. Overall, the marginal odds ratio was 2.0, indicating that the system worked two times better for males at avoiding lane deviation when disregarding level.

When running the Cochran-Maentel-Haenszel method to test for conditional independence of gender and lane departure, given the level of RDAS-HC I got a p-value of 7.338×10^{-6}, indicating that gender and lane departure are conditionally dependent given level of RDAS-HC. The common odds ratio was 2.09 with a 95% confidence interval of (1.51, 2.90), again telling us that males had higher odds of avoiding lane departure than females.

Table 1. Lane departure data by level and gender

	LD	no	yes
Level	Male		
Level0	female	10	50
	male	47	121
Level1	female	28	116
	male	77	121
Level2	female	24	84
	male	82	170

Table 2. Conditional and Marginal Odds Ratios

Level 0 OR	1.941748
Level 1 OR	2.638522
Level 2 OR	1.689189
Marg. OR	2.016129

Problem 2

In problem two I used Data1.dat. In this problem we used a logit model to further investigate the effects of the variables on lane departure. I fit a saturated logit model and then ran a backwards step function to select a model. The selected model included Level, Male, Night, Wet, Level*Wet, and Male*Night. The summary of the model (**Table 3**) and the standardized Pearson residuals (**Table 4**) are shown below. When we observe the residuals we can se that when it's dry, males have slightly worse performance than females. When it's wet females have worse performance than males. Overall, for both genders and regardless of the light condition, performance was slightly worse when the road was wet. The level of RDAS-HC does not appear to make much difference based on these residuals.

When we compare this chosen model to the main effects logit model, we see that the main effects logit model and our chosen model both fit well, as seen by having high p-values when testing for fit. The main effects model implies homogenous association, and the main effects model fits well, so we can assume homogenous association between RDAS-HC level and lane departure.

Next I used the likelihood ratio to test for conditional independence of RDAS-HC and response, given the other variables. I tested the null hypothesis of whether all beta's in the model equal zero, versus the alternative that at least one does not equal zero. I got a p-value of 0.92, so we can accept the null hypothesis that all beta's equal zero (**Table 5**). So, we can say that there is conditional independence between RDAS-HC level and the response, given the other variables. We can also use a Wald test for conditional independence. The Wald test for conditional independence is given in the summary of the logistic regression (**Table 3**). When I test the model given in part b, we see that the Wald p-value for level 0 vs. level 2 is 0.46, and the p-value for level 1 vs. level 2 is 0.32, suggesting that RDAS-HC level and response are conditionally independent, given the other variables. Finally, I constructed a ROC curve for the chosen model (**Figure 1**). When we look at the ROC curve, we can see that the area below the curve equals 0.563, implying that the predictive power of the chosen model is moderate.

Table 3. Output for chosen model

```
Call:
glm(formula = cbind(LDno, LDyes) ~ Level + Male + Night + Wet +
  Level:Wet + Male:Night, family = binomial, data = Tab1.logit)

Deviance Residuals:
   Min       1Q   Median       3Q      Max
-1.4531  -0.5223   0.1256   0.4618   0.8145

Coefficients:
                          Estimate  Std. Error  z value  Pr(>|z|)
(Intercept)               -0.60067   0.22306  -2.693   0.00708 **
LevelLevel0               -0.21770   0.29741  -0.732   0.46418
LevelLevel1               -0.30498   0.30652  -0.995   0.31976
Malefemale                -1.00987   0.24562  -4.111   3.93e-05 ***
Nightdaylight             -0.17710   0.17196  -1.030   0.30304
Wetdry                     0.02056   0.24852   0.083   0.93407
LevelLevel0:Wetdry        -0.06822   0.39635  -0.172   0.86335
LevelLevel1:Wetdry         0.65274   0.36628   1.782   0.07474 .
Malefemale:Nightdaylight   0.50414   0.33478   1.506   0.13210
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for binomial family taken to be 1)

    Null deviance: 40.4088  on 23  degrees of freedom
Residual deviance:  8.0854  on 15  degrees of freedom
AIC: 113.16
```

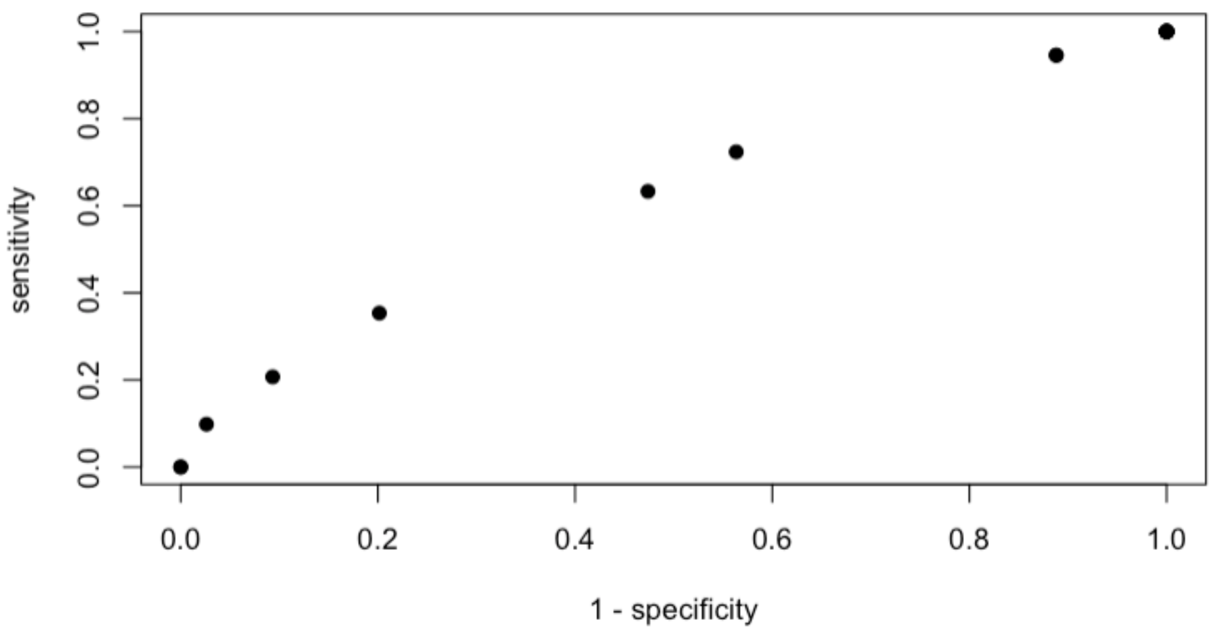
Table 4. Standardized Pearson residuals for chosen model

Level	Male	Night	Wet	LDno	LDyes	pear.std.indp
Level0	female	daylight	dry	3	12	0.278677741
Level1	female	daylight	dry	12	36	-0.761272477
Level2	female	daylight	dry	9	27	0.524126061
Level0	male	daylight	dry	12	30	0.506907882
Level1	male	daylight	dry	27	39	0.234589872
Level2	male	daylight	dry	24	60	-0.976486187
Level0	female	night	dry	2	13	0.005876361
Level1	female	night	dry	11	37	0.116541650
Level2	female	night	dry	8	28	1.021244273
Level0	male	night	dry	11	31	-0.679914705
Level1	male	night	dry	30	36	0.286298561
Level2	male	night	dry	30	54	-0.050801085
Level0	female	daylight	wet	2	13	-0.543406178
Level1	female	daylight	wet	4	20	-0.045038928
Level2	female	daylight	wet	5	13	0.713151911
Level0	male	daylight	wet	9	33	-1.107782324
Level1	male	daylight	wet	10	23	0.849908737
Level2	male	daylight	wet	15	27	0.802920909
Level0	female	night	wet	3	12	0.754206584
Level1	female	night	wet	1	23	-1.460132481
Level2	female	night	wet	2	16	-0.708773228
Level0	male	night	wet	15	27	1.005726343
Level1	male	night	wet	10	23	0.252372018
Level2	male	night	wet	13	29	-0.837718234

Table 5. Comparison of Likelihood ratio test and AIC for main effects model vs. the chosen model

	G2	X2	df	pval	AIC
Main effects	14.6774	13.5903	18	0.6840	113.7505
Model from part b	8.0854	7.6329	15	0.9203	113.1584

Figure 1. ROC curve for chosen model



Problems 3 & 4

For problems three and four I used Data2.dat. In these problems we treated Distraction as the response variable and Level and Male as the variables, and we looked at Distraction as both a nominal response variable and an ordinal response variable. First, I fitted a multinomial regression with Level and Male with a nominal response and compared the main effects model and the interaction model to see if the interaction term was needed (**Table 6**). When we contrast the multinomial regression with and without an interaction term for level and male, we can see that the anova test gives a p-value of 0, suggesting a significant difference between these models. We can see that the model with interaction has a lower AIC score, suggesting that it is a better model, so we would assume that we need the interaction term. Based on this model I then calculate the fitted counts for each combination of the levels (**Table 7**).

Next, I fitted a cumulative logit model in order to create a model with an ordinal response variable. Again, I compared the main effects model to the model with an interaction term. I could not get the regression to work using the vglm function, but I was able to do it using polr and lrm functions. Similar to the above, the AIC is lower for the interaction model (AIC = 3146.725) than for the main effects model (AIC = 3226.115), implying that the interaction model is the better model. Using the interaction model I was able to calculate (Distraction ≤ 0) for a male using level 1 and 2 (**Table 8**). For a male using level 1, the probability that Distraction is less than or equal to zero is 0.481 and for a male using level 2, the probability that Distraction is less than or equal zero is 0.399.

Table 6. ANOVA comparison of main effects model and interaction model

Model	Resid. df	Resid. Dev	Test	Df	LR stat.	Pr(Chi)
Level + Male	696	3069.723	NA	NA	NA	
Level * Male	684	2884.582	1 vs 2	12	185.1417	0

Table 7. Estimated counts from interaction model

Level	Male	-3	-2	-1	0	1	2	3
Level0	female	0	12	12	12	24	0	0
Level1	female	36	54	18	18	0	18	0
Level2	female	18	54	18	0	18	0	0
Level0	male	12	72	36	36	12	0	0
Level1	male	0	54	36	18	36	36	18
Level2	male	0	54	36	36	54	54	18

Table 8. Probabilities from cumulative model for ordinal response variable

Level	Male	y>=-2	y>=-1	y>=0	y>=1	y>=2	y>=3
1	1	0.9549510	0.6896339	0.5189977	0.3740970	0.1880334	0.04632008
2	1	0.9673327	0.7563305	0.6011562	0.4550163	0.2444229	0.06353658

Problem 5

In problem five I used Data3.dat. In this section we fitted log linear models to further analyze relationships between variables. First, I fitted a log linear model with the main effects of Distraction and LD and the interaction terms between Level, Male, Night, and Wet and compared it to a main effects model and a model chosen by automatic backward selection (**Table 9**), which ended up to be (LD, Distraction, Level*Male, Level*Wet). When we compare the model with the 4-way interaction term, a model found using backwards selection form the 4-way interaction term model, and the main effects model, we find that they all fit fine but the model with the best fit is model found with backwards selection (which does not have the 4-way interaction term). Additionally, the Standardized Pearson residuals for all three of these models are acceptable (**Table 10**). So, when comparing these models, we can say that we can drop the interaction term and it should not significantly effect the accuracy when comparing these three models to each other.

Next, I performed a automatic forward selection from the 4-way interaction model given at the beginning of this problem. The chosen model was (Distraction*LD*Male, Level*Male*Night*Wet). The association graph for this model can be seen below (**Figure 2**). When we consider whether distraction and LD are conditionally independent given the other variables, we can see from the association graph that they are not conditionally independent, as there is a connection between LD and distraction, they are interacting with each other. However, it is possible that the LD*Distraction interactions are not necessary to keep in the model. When I tested this model found with forward selection against the same model with all LD and distraction interaction removed I found that both models fit well, and the standardized Pearson residuals for the non-interaction model are fine (**Table 11**), so we could chose that model instead.

We can also consider whether there is an equivalent logit model to the log linear model we chose above. The model chosen before is (LD, Distraction, Level*Male, Level*Wet) which would not be equivalent to either the main effects logit model of the two-way interaction logit model. In order to be equivalent to the main effects logit model, the log linear model would need a five-way interaction term. Similarly, this log linear model is missing terms that would be necessary for the two-way interaction logit model.

Finally I considered the fit for the model chosen above: (LD, Distraction, Level*Male, Level*Wet) (**Table 12**). The model from part b have a p-value of 0.9999 when testing for fit against the saturated model, so this suggests a good fit. However, when we look at the standardized Pearson residuals, the minimum is negative infinity suggesting a poor fit. Because of this residual, I would not be able to say confidently that this model has a good fit.

Table 9. Likelihood ratio test and AIC for three models

	G2	X2	df	pval	AIC
4-way	178.1211	185.1291	185	0.6283073	906.3005
Backward fit	179.9114	187.0297	200	0.8429244	878.0908
Main effects	209.4974	222.4027	203	0.3623856	901.6768

Table 10. Standardized Pearson residual for three models

4-way Interaction

Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
-2.193374	-0.707752	-0.185979	0.003447	0.524358	3.311428

Main Effects

Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
-2.172732	-0.671382	-0.248855	0.001577	0.499560	3.693678

Backward Selection

Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
-2.0564	-0.6117	-0.1839	0.0026	0.4999	3.3607

Figure 2. Association graph for model (LD, Distraction, Level*Male, Level*Wet)

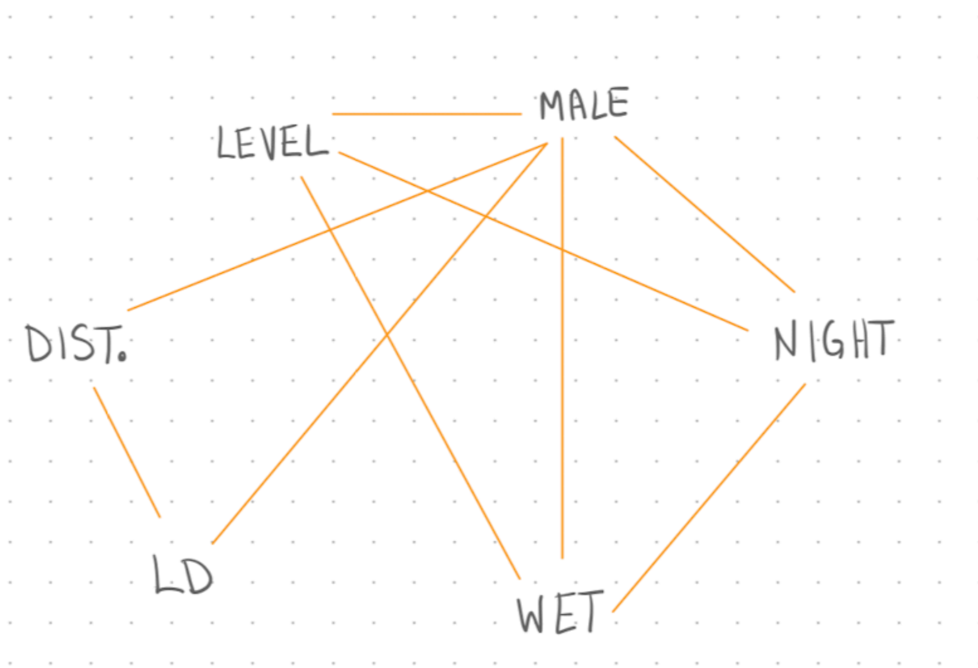


Table 11. Comparison of backward fit with and without LD*Distraction interaction

	G2	X2	df	pval	AIC
Backward fit without L*D int.	151.1845	152.7480	179	0.9355263	891.3639
Backward fit with L*D int.	179.9114	187.0297	200	0.8429244	878.0908

Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
-2.155422	-0.604187	-0.081850	0.008573	0.496194	3.057272

Table 12. Comparison of forward fit model to other considered models

	G2	X2	df	pval	AIC
Forward fit (b)	109.7448	107.5428	168	0.99984	871.9242
4-way mod	178.1211	185.1292	185	0.62831	906.3005
Backward fit	179.9114	187.0297	200	0.84292	878.0908
Main effects	209.4974	222.4027	203	0.36239	901.6768

Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
-Inf	-0.53063	-0.05556	-Inf	0.40394	2.56425

APPENDIX - ALL R CODE

Problem 1

Use the following codes to obtain a table.

```
Data1 <- read.table("Data1.dat",header = TRUE)

Data1$Level[Data1$Level==0] <- "Level0"
Data1$Level[Data1$Level==1] <- "Level1"
Data1$Level[Data1$Level==2] <- "Level2"
Data1$Male[Data1$Male==0] <- "female"
Data1$Male[Data1$Male==1] <- "male"
Data1$Night[Data1$Night==0] <- "daylight"
Data1$Night[Data1$Night==1] <- "night"
Data1$Wet[Data1$Wet==0] <- "dry"
Data1$Wet[Data1$Wet==1] <- "wet"
Data1$LD[Data1$LD==0] <- "no"
Data1$LD[Data1$LD==1] <- "yes"

Tab1 <- aggregate(Data1$Count,
           by=list(Level=Data1$Level, Male=Data1$Male, LD=Data1$LD),
           FUN=sum)
names(Tab1)[4]="Count"
Tab1

##     Level   Male  LD Count
## 1  Level0 female  no    10
## 2  Level1 female  no    28
## 3  Level2 female  no    24
## 4  Level0   male  no    47
## 5  Level1   male  no    77
## 6  Level2   male  no    82
## 7  Level0 female yes    50
## 8  Level1 female yes   116
## 9  Level2 female yes    84
```

```
## 10 Level0   male yes   121
## 11 Level1   male yes   121
## 12 Level2   male yes   170
```

Problem 1 Part a

(a) Reorganize the data to get a 3 × 2 × 2 contingency table containing **Level**, **Male**, and **LD** and present the table using functions xtabs and ftable. [3 pts]

```
Tab1.2 <- ftable(xtabs(Count~Level + Male + LD, data=Tab1),
       col.vars="LD")
Tab1.2

##               LD  no yes
## Level  Male
## Level0 female     10  50
##        male       47 121
## Level1 female     28 116
##        male       77 121
## Level2 female     24  84
##        male       82 170
```

Problem 1 Part b

(b) Calculate the observed conditional odds ratios giving the effectiveness of reducing lane departure occurrence for **males** compared to **females** given each level of RDAS-HC system and compare the conditional odds ratios to the marginal odds ratio between **Male** and **LD**. Present these using a nice table (Hint: you may need to use the tapply function). [7 pts]

For level 0, males were 1.9 times less likely to have lane deviation than females, for level 1 males were 2.6 times less likely to have lane deviation than females, and for level 2 males were 1.7 times less likely to have lane deviation. Overall, the marginal odds ratio was 2.0, indicating that the system worked two times better for males at avoiding lane deviation when disregarding level.

```
or <- function(x) x[1,1]*x[2,2]/x[1,2]/x[2,1]
Level0.OR <- 1/round(or(Tab1.2[1:2, ]), 3)
Level1.OR <- 1/round(or(Tab1.2[3:4, ]), 3)
Level2.OR <- 1/round(or(Tab1.2[5:6, ]), 3)
Marg.OR <- (47+77+82)*(50+116+84)/(121+121+170)/(10+28+24)
```

```
ors <- data.frame(rbind(Level0.OR,Level1.OR,Level2.OR,Marg.OR))
ors

##           rbind.Level0.OR..Level1.OR..Level2.OR..Marg.OR.
## Level0.OR                                        1.941748
## Level1.OR                                        2.638522
## Level2.OR                                        1.689189
## Marg.OR                                          2.016129
```

Problem 1 Parts c & d

(c) Use the Cochran-Maentel-Haenszel method to test for conditional independence of gender and lane departure, given the level of RDAS-HC. [5 pts]

(d) Give a 95% confidence interval for the common odds ratio based on the CMH method. [5 pts]

The p-value of the CMH test was 7.338x10^-6, indicating that gender and lane departure are conditionlly dependent given level of RDAS-HC. The common odds ratio was 2.09 with a 95% CI of (1.51, 2.90), again telling us that males had higher odds of avoiding lane departure than femles.

```
data <- expand.grid(Response = c("LD0", "LD1"),
                    Gender = c("Male","Female"),
                    Level = c(0,1,2))
data$count <- c(47,121,10,50,77,121,28,116,82,170,24,84)
data <- data[order(data$Response), ]
tab1.3 <- tapply(data$count, data[,c(2,1,3)],sum)
tab1.3

## , , Level = 0
##
##         Response
## Gender   LD0 LD1
##   Male    47 121
##   Female  10  50
##
## , , Level = 1
##
##         Response
```

```
## Gender   LD0 LD1
##   Male    77 121
##   Female  28 116
##
## , , Level = 2
##
##         Response
## Gender   LD0 LD1
##   Male    82 170
##   Female  24  84
```

```
cmh.test <- mantelhaen.test(tab1.3, correct=F)
cmh.test
```

```
##
##  Mantel-Haenszel chi-squared test without continuity correction
##
## data:  tab1.3
## Mantel-Haenszel X-squared = 20.103, df = 1, p-value = 7.338e-06
## alternative hypothesis: true common odds ratio is not equal to 1
## 95 percent confidence interval:
##  1.508708 2.903042
## sample estimates:
## common odds ratio
##          2.092808
```

Problem 2

Use data1.dat and treat **LD** as the response variable and other variables as predictors. **(Please just use data1.dat and do not run the codes in Problem 1)**.

Problem 2 Part a

(a) Reorganize the data so that the resulting data frame contains two columns that contain the counts of **LD yes** and **LD no**, respectively. [3 pts]

```
Tab1.logit <- data.frame(expand.grid(Level=c("Level0","Level1","Level2"),
                                     Male=c("female","male"),
                                     Night=c("daylight","night"),
Wet=c("dry","wet")),
                         LDno=Data1$Count[1:24],
                         LDyes=Data1$Count[25:48])
Tab1.logit

##     Level   Male    Night Wet LDno LDyes
## 1  Level0 female daylight dry    3    12
## 2  Level1 female daylight dry   12    36
## 3  Level2 female daylight dry    9    27
## 4  Level0   male daylight dry   12    30
## 5  Level1   male daylight dry   27    39
## 6  Level2   male daylight dry   24    60
## 7  Level0 female    night dry    2    13
## 8  Level1 female    night dry   11    37
## 9  Level2 female    night dry    8    28
## 10 Level0   male    night dry   11    31
## 11 Level1   male    night dry   30    36
## 12 Level2   male    night dry   30    54
## 13 Level0 female daylight wet    2    13
## 14 Level1 female daylight wet    4    20
## 15 Level2 female daylight wet    5    13
## 16 Level0   male daylight wet    9    33
## 17 Level1   male daylight wet   10    23
## 18 Level2   male daylight wet   15    27
## 19 Level0 female    night wet    3    12
## 20 Level1 female    night wet    1    23
```

```
## 21 Level2 female    night wet    2    16
## 22 Level0   male    night wet   15    27
## 23 Level1   male    night wet   10    23
## 24 Level2   male    night wet   13    29
```

Problem 2 Part b

(b) Fit a saturated logit model and use step function to select a model. [7 pts]

The backward selection chose a model including Level, Male, Night, Wet, LevelxWet, and MalexNight.

```
options(contrasts=c("contr.SAS","contr.poly"))
logit.sat <- glm(cbind(LDno,LDyes) ~.^4, data=Tab1.logit, family=binomial)
logit.step <- step(logit.sat, direction = "backward")

## Start:  AIC=135.07
## cbind(LDno, LDyes) ~ (Level + Male + Night + Wet)^4
##
##                        Df Deviance    AIC
## - Level:Male:Night:Wet  2  0.44649 131.52
## <none>                     0.00000 135.07
##
## Step:  AIC=131.52
## cbind(LDno, LDyes) ~ Level + Male + Night + Wet + Level:Male +
##     Level:Night + Level:Wet + Male:Night + Male:Wet + Night:Wet +
##     Level:Male:Night + Level:Male:Wet + Level:Night:Wet + Male:Night:Wet
##
##                    Df Deviance    AIC
## - Level:Male:Night  2   0.6667 127.74
## - Level:Male:Wet    2   0.7222 127.80
## - Male:Night:Wet    1   0.9558 130.03
## - Level:Night:Wet   2   4.2546 131.33
## <none>                  0.4465 131.52
##
## Step:  AIC=127.74
## cbind(LDno, LDyes) ~ Level + Male + Night + Wet + Level:Male +
##     Level:Night + Level:Wet + Male:Night + Male:Wet + Night:Wet +
```

```
##     Level:Male:Wet + Level:Night:Wet + Male:Night:Wet
##
##                   Df Deviance    AIC
## - Level:Male:Wet   2   0.9525 124.03
## - Male:Night:Wet   1   1.0649 126.14
## - Level:Night:Wet  2   4.4562 127.53
## <none>                 0.6667 127.74
##
## Step:  AIC=124.03
## cbind(LDno, LDyes) ~ Level + Male + Night + Wet + Level:Male +
##     Level:Night + Level:Wet + Male:Night + Male:Wet + Night:Wet +
##     Level:Night:Wet + Male:Night:Wet
##
##                   Df Deviance    AIC
## - Level:Male       2   2.8294 121.90
## - Male:Night:Wet   1   1.3042 122.38
## - Level:Night:Wet  2   4.6284 123.70
## <none>                 0.9525 124.03
##
## Step:  AIC=121.9
## cbind(LDno, LDyes) ~ Level + Male + Night + Wet + Level:Night +
##     Level:Wet + Male:Night + Male:Wet + Night:Wet + Level:Night:Wet +
##     Male:Night:Wet
##
##                   Df Deviance    AIC
## - Male:Night:Wet   1   3.1464 120.22
## - Level:Night:Wet  2   6.4905 121.56
## <none>                 2.8294 121.90
##
## Step:  AIC=120.22
## cbind(LDno, LDyes) ~ Level + Male + Night + Wet + Level:Night +
##     Level:Wet + Male:Night + Male:Wet + Night:Wet + Level:Night:Wet
##
##                   Df Deviance    AIC
## - Male:Wet         1   3.6694 118.74
```

```
## - Level:Night:Wet  2   6.9670 120.04
## <none>                 3.1464 120.22
## - Male:Night       1   5.3702 120.44
##
## Step:  AIC=118.74
## cbind(LDno, LDyes) ~ Level + Male + Night + Wet + Level:Night +
##     Level:Wet + Male:Night + Night:Wet + Level:Night:Wet
##
##                   Df Deviance    AIC
## - Level:Night:Wet  2   7.4572 118.53
## <none>                 3.6694 118.74
## - Male:Night       1   5.8430 118.92
##
## Step:  AIC=118.53
## cbind(LDno, LDyes) ~ Level + Male + Night + Wet + Level:Night +
##     Level:Wet + Male:Night + Night:Wet
##
##               Df Deviance    AIC
## - Level:Night  2   7.9014 114.97
## - Night:Wet    1   7.7431 116.82
## <none>             7.4572 118.53
## - Male:Night   1   9.5818 118.66
## - Level:Wet    2  11.8194 118.89
##
## Step:  AIC=114.97
## cbind(LDno, LDyes) ~ Level + Male + Night + Wet + Level:Wet +
##     Male:Night + Night:Wet
##
##              Df Deviance    AIC
## - Night:Wet   1   8.0854 113.16
## <none>            7.9014 114.97
## - Male:Night  1  10.2291 115.30
## - Level:Wet   2  12.2413 115.31
##
## Step:  AIC=113.16
```

```
## cbind(LDno, LDyes) ~ Level + Male + Night + Wet + Level:Wet +
##     Male:Night
##
##              Df Deviance    AIC
## <none>            8.0854 113.16
## - Male:Night  1  10.3661 113.44
## - Level:Wet   2  12.4084 113.48
```

```
summary(logit.step)
```

```
##
## Call:
## glm(formula = cbind(LDno, LDyes) ~ Level + Male + Night + Wet +
##     Level:Wet + Male:Night, family = binomial, data = Tab1.logit)
##
## Deviance Residuals:
##     Min       1Q   Median       3Q      Max
## -1.4531  -0.5223   0.1256   0.4618   0.8145
##
## Coefficients:
##                            Estimate Std. Error z value Pr(>|z|)
## (Intercept)                -0.60067    0.22306  -2.693  0.00708 **
## LevelLevel0                -0.21770    0.29741  -0.732  0.46418
## LevelLevel1                -0.30498    0.30652  -0.995  0.31976
## Malefemale                 -1.00987    0.24562  -4.111 3.93e-05 ***
## Nightdaylight              -0.17710    0.17196  -1.030  0.30304
## Wetdry                      0.02056    0.24852   0.083  0.93407
## LevelLevel0:Wetdry         -0.06822    0.39635  -0.172  0.86335
## LevelLevel1:Wetdry          0.65274    0.36628   1.782  0.07474 .
## Malefemale:Nightdaylight  0.50414    0.33478   1.506  0.13210
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 40.4088  on 23  degrees of freedom
```

```
## Residual deviance:  8.0854  on 15  degrees of freedom
## AIC: 113.16
##
## Number of Fisher Scoring iterations: 4
```

Problem 2 Part c

(c) Obtain the standardized Pearson residuals from the model selected in part (b). What do you conclude? [5 pts]

When it's dry, males have slightly worse performance than females. When it's wet females have worse performance than males. Overall, for both genders and regardless of the light condition, performance was slightly worse when the road was wet. The level of RDAS-HC does not appear to make much difference based on these residuals.

```
pear.indep <- resid(logit.step, type="pearson")
pear.std.indp <- pear.indep/sqrt(1-lm.influence(logit.step)$hat)

residtab <- data.frame(cbind(Tab1.logit),pear.std.indp)
residtab
```

```
##     Level   Male    Night Wet LDno LDyes pear.std.indp
## 1  Level0 female daylight dry    3    12   0.278677741
## 2  Level1 female daylight dry   12    36  -0.761272477
## 3  Level2 female daylight dry    9    27   0.524126061
## 4  Level0   male daylight dry   12    30   0.506907882
## 5  Level1   male daylight dry   27    39   0.234589872
## 6  Level2   male daylight dry   24    60  -0.976486187
## 7  Level0 female    night dry    2    13   0.005876361
## 8  Level1 female    night dry   11    37   0.116541650
## 9  Level2 female    night dry    8    28   1.021244273
## 10 Level0   male    night dry   11    31  -0.679914705
## 11 Level1   male    night dry   30    36   0.286298561
## 12 Level2   male    night dry   30    54  -0.050801085
## 13 Level0 female daylight wet    2    13  -0.543406178
## 14 Level1 female daylight wet    4    20  -0.045038928
## 15 Level2 female daylight wet    5    13   0.713151911
## 16 Level0   male daylight wet    9    33  -1.107782324
## 17 Level1   male daylight wet   10    23   0.849908737
```

```
## 18 Level2   male daylight wet   15    27   0.802920909
## 19 Level0 female    night wet    3    12   0.754206584
## 20 Level1 female    night wet    1    23  -1.460132481
## 21 Level2 female    night wet    2    16  -0.708773228
## 22 Level0   male    night wet   15    27   1.005726343
## 23 Level1   male    night wet   10    23   0.252372018
## 24 Level2   male    night wet   13    29  -0.837718234
```

Problem 2 Part d

(d) Test whether the association between RDAS-HC level and lane departure is homogeneous, given the other variables. (Hint: which logit model implies the homogeneous association). [5 pts]

Main effects model for logit implies homogeneous association. The main effects model fits well, so we can assume homogeneous association.

```
fit.maineff <- glm(cbind(LDno,LDyes) ~., data=Tab1.logit, family=binomial)
models <- list(fit.maineff,logit.step)
mod.G2 <- sapply(models,function(x)x$deviance)
mod.X2 <- sapply(models,function(x) sum(residuals(x,type="pearson")^2))
mod.df <- sapply(models,function(x) x$df.resid)
mod.pval <- pchisq(mod.G2,mod.df,lower=F)
mod.AIC <- sapply(models,function(x) x$aic)
lackFit <-
data.frame(G2=mod.G2,X2=mod.X2,df=mod.df,pval=mod.pval,AIC=mod.AIC)
rownames(lackFit) <- c("Main effects","Model from part b")
round(lackFit, 4)

##                         G2      X2 df   pval      AIC
## Main effects      14.6774 13.5903 18 0.6840 113.7505
## Model from part b  8.0854  7.6329 15 0.9203 113.1584
```

Problem 2 Part e

(e) Use likelihood ratio test to test for conditional independence of RDAS-HC and response, given the other variables. Be sure to clearly state the model and hypothesis you are testing. [5 pts]

I am testing the null hypothesis of whether all beta’s in the model equal zero, versus the alternative that at least one does not equal zero. We get a p-value of 0.92, so we can accept the null hypothesis that all beta's equal zero. So, we can say that there is conditional independence between RDAS-HC level and the response, given the other variables.

```
logit.step$deviance

## [1] 8.085368

pchisq(logit.step$deviance, logit.step$df.residual, lower=F)

## [1] 0.9203146
```

Problem 2 BONUS

Bonus: Use the Wald test to test for conditional independence of RDAS-HC and response, given the other variables. [5 pts]

The wald test for conditional independence is given in the summary of the logistic regression. When I test the model given in part b, we see that the wald p-value for level 0 vs. level 2 is 0.46, and the p-value for level 1 vs. level 2 is 0.32, suggesting that RDAS-HC level and response are conditionally independent, given the other variables.

```
summary(logit.step)

##
## Call:
## glm(formula = cbind(LDno, LDyes) ~ Level + Male + Night + Wet +
##     Level:Wet + Male:Night, family = binomial, data = Tab1.logit)
##
## Deviance Residuals:
##     Min       1Q   Median       3Q      Max
## -1.4531  -0.5223   0.1256   0.4618   0.8145
##
## Coefficients:
##                         Estimate Std. Error z value Pr(>|z|)
## (Intercept)             -0.60067    0.22306  -2.693  0.00708 **
## LevelLevel0             -0.21770    0.29741  -0.732  0.46418
## LevelLevel1             -0.30498    0.30652  -0.995  0.31976
## Malefemale              -1.00987    0.24562  -4.111 3.93e-05 ***
## Nightdaylight           -0.17710    0.17196  -1.030  0.30304
## Wetdry                   0.02056    0.24852   0.083  0.93407
## LevelLevel0:Wetdry      -0.06822    0.39635  -0.172  0.86335
```

```
## LevelLevel1:Wetdry        0.65274    0.36628   1.782  0.07474 .
## Malefemale:Nightdaylight  0.50414    0.33478   1.506  0.13210
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 40.4088  on 23  degrees of freedom
## Residual deviance:  8.0854  on 15  degrees of freedom
## AIC: 113.16
##
## Number of Fisher Scoring iterations: 4
```

Problem 2 Part f

(f) Draw the ROC curve for the model selected in Part (b). You may use or modify the following codes to create a suitable data structure. Let Tab2 be the data frame you created in (a) and the column names of LD Yes and No be Yes and No. I assume that they are columns 5 and 6. If not, you can modify the codes.

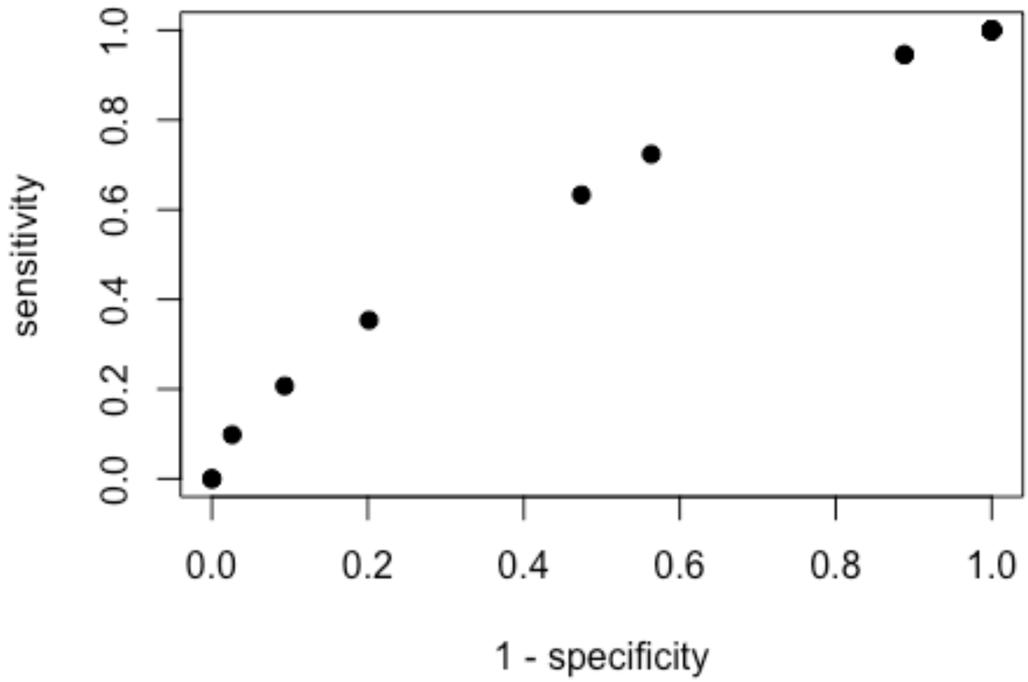
```
#Tab3.1=(Tab2[rep(seq_len(nrow(Tab2)), Tab2$No),])[,-c(5,6)]
#Tab3.1$y = 0
#Tab3.2=(Tab2[rep(seq_len(nrow(Tab2)), Tab2$Yes),])[,-c(5,6)]
#Tab3.2$y = 1
#Tab3=rbind(Tab3.1, Tab3.2)
```

Fit Tab3 with the model you selected in Part 2(b) and produce the ROC curve.

When we look at the ROC curve, we can see that the area below the curve equals 0.563, implying that the predictive power of the the model selected in part b is moderate.

```
Tab2.1=(Tab1.logit[rep(seq_len(nrow(Tab1.logit)), Tab1.logit$LDno),])[,-
c(5,6)]
Tab2.1$y = 0
Tab2.2=(Tab1.logit[rep(seq_len(nrow(Tab1.logit)), Tab1.logit$LDyes),])[,-
c(5,6)]
Tab2.2$y = 1
Tabroc=rbind(Tab2.1, Tab2.2)
```

```r
fit.roc <- glm(y ~ Level + Male + Night + Wet + Level*Night + Level*Wet +
Male*Night + Night*Wet + Level*Night*Wet,
               data=Tabroc, family=binomial)


pihat <- predict(fit.roc,type="response")

pi0 <- seq(0.05,0.95,by=.05)
fun <- function(x,y) ifelse(x>y,1,0)
sensfun <-function(ypred) sum(ypred[Tabroc$y==1]==1)/sum(Tabroc$y==1)
specfun <-function(ypred) sum(ypred[Tabroc$y==0]==0)/sum(Tabroc$y==0)
roc <- function(arg1) {
  yhat <- outer(arg1,pi0,fun)
  sens <- apply(yhat,2,sensfun)
  spec <- apply(yhat,2,specfun)
  data.frame(sens=sens,spec=spec)
}

x <- roc(pihat)
f <- approxfun(1-x$spec,x$sens)

## Warning in regularize.values(x, y, ties, missing(ties), na.rm = na.rm):
## collapsing to unique 'x' values

area <- integrate(f,0.05,0.95)$value
plot(1-x$spec,x$sens,xlab="1 - specificity",
       ylab="sensitivity",pch=19)
```




```
area <- round(area,3)
area

## [1] 0.563
```

Problem 3

Use data2.dat and treat **Distraction** as a **nominal** response variable and other variables as predictors.

```
Data2 <- read.table("Data2.dat",header = TRUE)
Data2$Level[Data2$Level==0] <- "Level0"
Data2$Level[Data2$Level==1] <- "Level1"
Data2$Level[Data2$Level==2] <- "Level2"
Data2$Male[Data2$Male==0] <- "female"
Data2$Male[Data2$Male==1] <- "male"
Data2$Distraction <- factor(Data2$Distraction)
```

Problem 3 Part a

(a) Fit a multinomial regression with variables **Level** and **Male**. Whether the interaction term should be included in the model? [5 pts]

When we contrast the multinomial regression with and without an interaction term for level and male, we can see that the anova test gives a p-value of 0, suggesting a significant difference between these models. We can see that the model with interaction has a lower AIC score, suggesting that it is a better model, so we would assume that we do need the interaction term.

```
library(nnet)
options(contrasts=c("contr.SAS","contr.poly"))
fit3.1 <- multinom(Distraction ~ Level+Male, data=Data2, weights=Count,
trace=F)
fit3.2 <- multinom(Distraction ~ Level*Male, data=Data2, weights=Count,
trace=F)

anova(fit3.1,fit3.2,test="Chisq")

##          Model Resid. df Resid. Dev   Test    Df LR stat. Pr(Chi)
## 1 Level + Male       696   3069.723           NA       NA      NA
## 2 Level * Male       684   2884.582 1 vs 2    12 185.1417       0

summary(fit3.1)

## Call:
## multinom(formula = Distraction ~ Level + Male, data = Data2,
```

```
##     weights = Count, trace = F)
##
## Coefficients:
##    (Intercept)  LevelLevel0 LevelLevel1 Malefemale
## -2    2.892726  -0.01454002  -0.5067074  -1.831864
## -1    2.334270   0.08603646  -0.4613879  -2.231431
## 0     1.992801   0.47462946  -0.4367863  -2.468634
## 1     2.636078  -0.49299962  -1.1493127  -2.280363
## 2     2.538180 -15.18848648  -0.3745379  -3.221335
## 3     1.583236 -15.97181030  -0.2955991 -18.343966
##
## Std. Errors:
##    (Intercept)  LevelLevel0 LevelLevel1   Malefemale
## -2   0.3671979 4.098241e-01   0.3269511 3.439120e-01
## -1   0.3814358 4.349797e-01   0.3578025 3.671915e-01
## 0    0.3940186 4.472914e-01   0.3849348 3.868827e-01
## 1    0.3767078 4.383756e-01   0.3641248 3.730108e-01
## 2    0.3825934 2.910143e+02   0.3662756 4.160699e-01
## 3    0.4290929 1.036539e-04   0.4637423 8.531718e-06
##
## Residual Deviance: 3069.723
## AIC: 3117.723
```

```
summary(fit3.2)
```

```
## Call:
## multinom(formula = Distraction ~ Level * Male, data = Data2,
##     weights = Count, trace = F)
##
## Coefficients:
##    (Intercept) LevelLevel0 LevelLevel1 Malefemale LevelLevel0:Malefemale
## -2    19.46875   -17.67699   10.066526  -18.37014               37.82354
## -1    19.06329   -17.96467   10.066527  -19.06328               39.20983
## 0     19.06328   -17.96467    9.373382  -37.66727               57.81381
## 1     19.46875   -19.46875    9.661063  -19.46875               41.40705
## 2     19.46876   -34.62784    9.661060  -39.49589               22.60725
```

```
## 3     18.37014   -34.79790   10.066532  -34.82301               23.49604
##    LevelLevel1:Malefemale
## -2            -10.759671
## -1            -10.759675
## 0               8.537452
## 1             -25.427054
## 2               9.672934
## 3             -15.687503
##
## Std. Errors:
##    (Intercept) LevelLevel0 LevelLevel1 Malefemale LevelLevel0:Malefemale
## -2    3.073375   3.0811157    43.88873  3.0803344           4.322947e+00
## -1    3.074181   3.0838211    43.88885  3.0861350           4.325839e+00
## 0     3.074941   3.0848304    43.88916 14.1728177           1.293256e+01
## 1     3.073369   3.0906043    43.88879  3.0853079           4.326073e+00
## 2     3.074309   3.7638396    43.88888 20.4849443           3.993286e-16
## 3     3.079400   0.5625311    43.88949  0.8140898           3.546901e-15
##    LevelLevel1:Malefemale
## -2           43.88931988
## -1           43.89010486
## 0            26.64842621
## 1           398.02225691
## 2            20.33532305
## 3             0.06575573
##
## Residual Deviance: 2884.582
## AIC: 2956.582
```

Problem 3 Part b

(b) Based on the model in part (a), calculate the fitted counts for each combination of levels of **Distraction**, **Level** and **Male** (Hint: follow the gator food example) [5 pts]

```
newdata <- expand.grid(Level=c("Level0","Level1","Level2"),
                       Male=c("female","male"))
marg.count <- as.vector(tapply(Data2$Count, list(Data2$Level,Data2$Male),
                               sum))
pred.count <- predict(fit3.2, newdata, type="probs")*marg.count
```

```r
pred.count <- round(pred.count, 3)
newdata <- cbind(newdata, pred.count)
newdata
```

```
##    Level   Male -3 -2 -1  0  1  2  3
## 1 Level0 female  0 12 12 12 24  0  0
## 2 Level1 female 36 54 18 18  0 18  0
## 3 Level2 female 18 54 18  0 18  0  0
## 4 Level0   male 12 72 36 36 12  0  0
## 5 Level1   male  0 54 36 18 36 36 18
## 6 Level2   male  0 54 36 36 54 54 18
```

Problem 4

Use data2.dat and treat **Distraction** as an **ordinal** response variable and other variables as predictors.

Problem 4 Part a

(a) Fit a cumulative logit model with variables **Level** and **Male**. Whether the interaction term should be included in the model? (Hint: you may use the function vglm with logit link) [5 pts]

I could not get the regression to work using vglm, but I was able to do it using polr and lrm. Similar to the above, the AIC is lower for the interaction model, implying that it is the better model.

```
Data2 <- read.table("Data2.dat",header = TRUE)
Data2$Distraction <- factor(Data2$Distraction)

library(rms)

## Loading required package: Hmisc

##
## Attaching package: 'Hmisc'

## The following objects are masked from 'package:base':
##
##     format.pval, units

## Warning in !is.null(rmarkdown::metadata$output) &&
rmarkdown::metadata$output
## %in% : 'length(x) = 2 > 1' in coercion to 'logical(1)'

fit3.3 <- lrm(Distraction ~ Level+Male, data=Data2, weights=Count)

## Warning in lrm(Distraction ~ Level + Male, data = Data2, weights = Count):
## currently weights are ignored in model validation and bootstrapping lrm
fits

fit3.3

## Logistic Regression Model
##
```

```
## lrm(formula = Distraction ~ Level + Male, data = Data2, weights = Count)
##
##
## Frequencies of Responses
##
## -3 -2 -1  0  1  2  3
## 12 24 24 20 20 12  8
##
##
## Sum of Weights by Response Category
##
##   0   1   2   3   4   5   6
##  66 300 156 120 144 108  36
##
##                        Model Likelihood      Discrimination    Rank
Discrim.
##                              Ratio Test             Indexes
Indexes
## Obs            120    LR chi2     102.81      R2       0.108    C
0.641
## Sum of weights930    d.f.             2      R2(2,930)0.103    Dxy
0.282
## max |deriv| 1e-10    Pr(> chi2) <0.0001    R2(2,887.1)0.107    gamma
0.347
##                                              Brier    0.231    tau-a
0.228
##
##       Coef    S.E.   Wald Z Pr(>|Z|)
## y>=-2  1.5260 0.1698   8.99 <0.0001
## y>=-1 -0.7295 0.1433  -5.09 <0.0001
## y>=0  -1.4519 0.1480  -9.81 <0.0001
## y>=1  -2.0426 0.1548 -13.19 <0.0001
## y>=2  -2.9908 0.1712 -17.47 <0.0001
## y>=3  -4.5527 0.2275 -20.02 <0.0001
## Level  0.3343 0.0744   4.49 <0.0001
## Male   1.1937 0.1325   9.01 <0.0001
```

```
fit3.4 <- lrm(Distraction ~ Level*Male, data=Data2, weights=Count)

## Warning in lrm(Distraction ~ Level * Male, data = Data2, weights = Count):
## currently weights are ignored in model validation and bootstrapping lrm
fits

fit3.4

## Logistic Regression Model
##
## lrm(formula = Distraction ~ Level * Male, data = Data2, weights = Count)
##
##
## Frequencies of Responses
##
## -3 -2 -1  0  1  2  3
## 12 24 24 20 20 12  8
##
##
## Sum of Weights by Response Category
##
##   0   1   2   3   4   5   6
##  66 300 156 120 144 108  36
##
##                          Model Likelihood      Discrimination    Rank
Discrim.
##                                Ratio Test             Indexes
Indexes
## Obs             120     LR chi2     184.20      R2        0.185     C
0.676
## Sum of weights930    d.f.             3      R2(3,930)0.177    Dxy
0.352
## max |deriv| 2e-08    Pr(> chi2) <0.0001    R2(3,887.1)0.185    gamma
0.432
##                                                Brier    0.215    tau-a
0.284
##
##                Coef    S.E.   Wald Z Pr(>|Z|)
```

```
## y>=-2         2.9048 0.2263  12.83 <0.0001
## y>=-1         0.5366 0.1891   2.84 0.0045
## y>=0         -0.2386 0.1868  -1.28 0.2016
## y>=1         -0.8696 0.1884  -4.62 <0.0001
## y>=2         -1.8612 0.1991  -9.35 <0.0001
## y>=3         -3.4489 0.2484 -13.88 <0.0001
## Level        -0.7762 0.1422  -5.46 <0.0001
## Male         -0.4761 0.2202  -2.16 0.0306
## Level * Male  1.5245 0.1700   8.97 <0.0001
```

```
library(MASS)
fit3.7 <- polr(Distraction ~ Level+Male, data=Data2, weight=Count)
summary(fit3.7)
```

```
##
## Re-fitting to get Hessian
```

```
## Call:
## polr(formula = Distraction ~ Level + Male, data = Data2, weights = Count)
##
## Coefficients:
##        Value Std. Error t value
## Level 0.3343     0.0744   4.493
## Male  1.1937     0.1325   9.009
##
## Intercepts:
##       Value   Std. Error t value
## -3|-2 -1.5260  0.1697    -8.9895
## -2|-1  0.7295  0.1433     5.0903
## -1|0   1.4519  0.1480     9.8113
## 0|1    2.0426  0.1548    13.1937
## 1|2    2.9908  0.1712    17.4675
## 2|3    4.5527  0.2275    20.0160
##
## Residual Deviance: 3210.115
## AIC: 3226.115
```

```r
fit3.8 <- polr(Distraction ~ Level*Male, data=Data2, weight=Count)
summary(fit3.8)

##
## Re-fitting to get Hessian

## Call:
## polr(formula = Distraction ~ Level * Male, data = Data2, weights = Count)
##
## Coefficients:
##              Value Std. Error t value
## Level      -0.7762     0.1422  -5.459
## Male       -0.4761     0.2202  -2.162
## Level:Male  1.5245     0.1700   8.969
##
## Intercepts:
##       Value    Std. Error t value
## -3|-2  -2.9048   0.2263   -12.8340
## -2|-1  -0.5366   0.1891    -2.8380
## -1|0    0.2386   0.1868     1.2771
## 0|1     0.8696   0.1884     4.6168
## 1|2     1.8612   0.1991     9.3502
## 2|3     3.4488   0.2484    13.8825
##
## Residual Deviance: 3128.725
## AIC: 3146.725

#library(VGAM)
#fit3.5 = vglm(Distraction ~ Level+Male, family=cumulative(link =
"logitlink", parallel = FALSE), data=Data2)
#fit3.5
#g2.add.o = deviance(fit3.5)
#df.add.o = df.residual(fit3.5)
#1 - pchisq(g2.add.o, df.add.o)

#fit3.6 = vglm(Distraction ~ Level*Male, family=cumulative(link =
"logitlink", parallel = FALSE), data=Data2)
```

```
#g2.add = deviance(fit3.6)
#df.add = df.residual(fit3.6)
#1 - pchisq(g2.add, df.add)
```

Problem 4 Part b

(b) Based on the model in part (a), what are P (Distraction ≤ 0) for a male using level 1 and 2? [5 pts]

For a male using level 1, the probability that Distraction is less than or equal to zero is 0.481 and for a male using level 2, the probability that Distraction is less than or equal zero is 0.399.

```
newdata <- data.frame(expand.grid(Level=c(1,2), Male=1))
newdata <- cbind(newdata,predict(fit3.3,newdata,type="fitted"))
newdata

##   Level Male     y>=-2     y>=-1      y>=0      y>=1      y>=2       y>=3
## 1     1    1 0.9549510 0.6896339 0.5189977 0.3740970 0.1880334 0.04632008
## 2     2    1 0.9673327 0.7563305 0.6011562 0.4550163 0.2444229 0.06353658

1-0.5189977

## [1] 0.4810023

1-0.6011562

## [1] 0.3988438
```

Problem 5

Use data3.dat. Recall that **Distraction** and **LD** are response variables.

```
Data3 <- read.table("Data3.dat",header = TRUE)
Data3$Level <- factor(Data3$Level)
Data3$Male[Data3$Male==0] <- "female"
Data3$Male[Data3$Male==1] <- "male"
Data3$Night[Data3$Night==0] <- "daylight"
Data3$Night[Data3$Night==1] <- "night"
Data3$Wet[Data3$Wet==0] <- "dry"
Data3$Wet[Data3$Wet==1] <- "wet"
Data3$LD[Data3$LD==0] <- "no"
Data3$LD[Data3$LD==1] <- "yes"
Data3$Distraction <- factor(Data3$Distraction)
```

Problem 5 Part a

(a) Fit a log linear model including the main effects of **Distraction** and **LD** and the interaction terms between **Level**, **Male**, **Night**, and **Wet.** Could we drop the interaction term? [5 pts]

When we compare the model with the 4-way interaction term, a model found using backwards selection form the 4-way interaction term model, and the main effects model, we find that they all fit fine but the model with the best fit is model found with backwards selection (which does not have the 4-way interaction term). So, when comparing these models, we can say that we can drop the interaction term.

```
fit.4waymod <- glm(Count ~. + Level*Male*Night*Wet, data = Data3, family =
poisson)
fit.main <- glm(Count ~., data = Data3, family = poisson)
fit.backward4 <- step(fit.4waymod, direction = "backward")

## Start:  AIC=906.3
## Count ~ Level + Male + Night + Wet + LD + Distraction + Level *
##     Male * Night * Wet
##
##                        Df Deviance    AIC
## - Level:Male:Night:Wet  2   178.23 902.41
## <none>                      178.12 906.30
```

```
## - Distraction           6   281.13  997.31
## - LD                    1   290.84 1017.02
##
## Step:  AIC=902.41
## Count ~ Level + Male + Night + Wet + LD + Distraction + Level:Male +
##     Level:Night + Male:Night + Level:Wet + Male:Wet + Night:Wet +
##     Level:Male:Night + Level:Male:Wet + Level:Night:Wet + Male:Night:Wet
##
##                    Df Deviance     AIC
## - Level:Male:Night  2   178.24  898.42
## - Level:Night:Wet   2   178.27  898.45
## - Level:Male:Wet    2   178.28  898.46
## - Male:Night:Wet    1   178.33  900.51
## <none>                  178.23  902.41
## - Distraction       6   281.23  993.41
## - LD                1   291.15 1013.33
##
## Step:  AIC=898.42
## Count ~ Level + Male + Night + Wet + LD + Distraction + Level:Male +
##     Level:Night + Male:Night + Level:Wet + Male:Wet + Night:Wet +
##     Level:Male:Wet + Level:Night:Wet + Male:Night:Wet
##
##                   Df Deviance     AIC
## - Level:Night:Wet  2   178.29  894.47
## - Level:Male:Wet   2   178.30  894.48
## - Male:Night:Wet   1   178.36  896.54
## <none>                 178.24  898.42
## - Distraction      6   281.24  989.42
## - LD               1   291.18 1009.36
##
## Step:  AIC=894.47
## Count ~ Level + Male + Night + Wet + LD + Distraction + Level:Male +
##     Level:Night + Male:Night + Level:Wet + Male:Wet + Night:Wet +
##     Level:Male:Wet + Male:Night:Wet
##
```

```
##                   Df Deviance     AIC
## - Level:Male:Wet  2   178.34  890.52
## - Level:Night     2   178.42  890.60
## - Male:Night:Wet  1   178.43  892.61
## <none>                178.29  894.47
## - Distraction     6   281.26  985.44
## - LD              1   291.34 1005.52
##
## Step:  AIC=890.52
## Count ~ Level + Male + Night + Wet + LD + Distraction + Level:Male +
##     Level:Night + Male:Night + Level:Wet + Male:Wet + Night:Wet +
##     Male:Night:Wet
##
##                   Df Deviance     AIC
## - Level:Night     2   178.48  886.66
## - Male:Night:Wet  1   178.48  888.66
## <none>                178.34  890.52
## - Level:Male      2   190.74  898.92
## - Level:Wet       2   196.20  904.38
## - Distraction     6   281.27  981.45
## - LD              1   291.53 1001.71
##
## Step:  AIC=886.66
## Count ~ Level + Male + Night + Wet + LD + Distraction + Level:Male +
##     Male:Night + Level:Wet + Male:Wet + Night:Wet + Male:Night:Wet
##
##                   Df Deviance    AIC
## - Male:Night:Wet  1   178.62 884.80
## <none>                178.48 886.66
## - Level:Male      2   190.86 895.04
## - Level:Wet       2   196.32 900.50
## - Distraction     6   281.58 977.76
## - LD              1   291.89 998.07
##
## Step:  AIC=884.8
```

```
## Count ~ Level + Male + Night + Wet + LD + Distraction + Level:Male +
##     Male:Night + Level:Wet + Male:Wet + Night:Wet
##
##               Df Deviance    AIC
## - Night:Wet    1   178.63 882.81
## - Male:Wet     1   178.95 883.13
## - Male:Night   1   179.35 883.53
## <none>             178.62 884.80
## - Level:Male   2   190.97 893.14
## - Level:Wet    2   196.46 898.64
## - Distraction  6   281.65 975.83
## - LD           1   292.35 996.53
##
## Step:  AIC=882.81
## Count ~ Level + Male + Night + Wet + LD + Distraction + Level:Male +
##     Male:Night + Level:Wet + Male:Wet
##
##               Df Deviance    AIC
## - Male:Wet     1   178.96 881.14
## - Male:Night   1   179.35 881.53
## <none>             178.63 882.81
## - Level:Male   2   190.97 891.15
## - Level:Wet    2   196.48 896.66
## - Distraction  6   281.66 973.84
## - LD           1   292.39 994.57
##
## Step:  AIC=881.14
## Count ~ Level + Male + Night + Wet + LD + Distraction + Level:Male +
##     Male:Night + Level:Wet
##
##               Df Deviance    AIC
## - Male:Night   1   179.67 879.85
## <none>             178.96 881.14
## - Level:Male   2   191.01 889.19
## - Level:Wet    2   196.57 894.75
```

```
## - Distraction  6   281.86 972.04
## - LD           1   293.38 993.55
##
## Step:  AIC=879.85
## Count ~ Level + Male + Night + Wet + LD + Distraction + Level:Male +
##     Level:Wet
##
##                Df Deviance    AIC
## - Night        1   179.91 878.09
## <none>              179.67 879.85
## - Level:Male   2   191.70 887.88
## - Level:Wet    2   197.36 893.54
## - Distraction  6   282.34 970.52
## - LD           1   295.34 993.52
##
## Step:  AIC=878.09
## Count ~ Level + Male + Wet + LD + Distraction + Level:Male +
##     Level:Wet
##
##                Df Deviance    AIC
## <none>              179.91 878.09
## - Level:Male   2   191.95 886.13
## - Level:Wet    2   197.63 891.81
## - Distraction  6   282.54 968.72
## - LD           1   296.05 992.23
```

```
models <- list(fit.4waymod,fit.backward4,fit.main)
mod.G2 <- sapply(models,function(x)x$deviance)
mod.X2 <- sapply(models,function(x) sum(residuals(x,type="pearson")^2))
mod.df <- sapply(models,function(x) x$df.resid)
mod.pval <- pchisq(mod.G2,mod.df,lower=F)
mod.AIC <- sapply(models,function(x) x$aic)
lackFit <-
data.frame(G2=mod.G2,X2=mod.X2,df=mod.df,pval=mod.pval,AIC=mod.AIC)
rownames(lackFit) <- c("4-way mod","Backward fit","Main effects")
round(lackFit, 10)
```

```
##                        G2       X2  df      pval      AIC
## 4-way mod    178.1211 185.1291 185 0.6283073 906.3005
## Backward fit 179.9114 187.0297 200 0.8429244 878.0908
## Main effects 209.4974 222.4027 203 0.3623856 901.6768
```

```
res.4way <- resid(fit.4waymod, type = "pearson")/sqrt(1
-lm.influence(fit.4waymod)$hat)
res.main <- resid(fit.main, type = "pearson")/sqrt(1
              - lm.influence(fit.main)$hat)
res.backward4 <- resid(fit.backward4, type = "pearson")/sqrt(1
              - lm.influence(fit.backward4)$hat)
summary(res.4way)
```

```
##      Min.   1st Qu.    Median      Mean   3rd Qu.      Max.
## -2.193374 -0.707752 -0.185979  0.003447  0.524358  3.311428
```

```
summary(res.main)
```

```
##      Min.   1st Qu.    Median      Mean   3rd Qu.      Max.
## -2.172732 -0.671382 -0.248855  0.001577  0.499560  3.693678
```

```
summary(res.backward4)
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## -2.0564 -0.6117 -0.1839  0.0026  0.4999  3.3607
```

Problem 5 Part b

(b) Start with the model in Part (a) and carry out a forward selection. [5 pts]

```
fit.forward4 <- step(fit.4waymod, scope = list(upper = ~.^6), direction =
"forward")
```

```
## Start:  AIC=906.3
## Count ~ Level + Male + Night + Wet + LD + Distraction + Level *
##     Male * Night * Wet
##
##                        Df Deviance    AIC
## + LD:Distraction        6   148.28 888.46
## + Male:Distraction      5   158.27 896.45
## + Level:Distraction    10   152.27 900.45
## + Male:LD               1   171.81 901.98
```

```
## <none>                     178.12 906.30
## + Night:LD           1   176.75 906.93
## + Level:LD           2   175.96 908.14
## + Wet:LD             1   178.02 908.20
## + Wet:Distraction    6   176.49 916.67
## + Night:Distraction  6   177.46 917.63
##
## Step:  AIC=888.46
## Count ~ Level + Male + Night + Wet + LD + Distraction + Level:Male +
##     Level:Night + Male:Night + Level:Wet + Male:Wet + Night:Wet +
##     LD:Distraction + Level:Male:Night + Level:Male:Wet + Level:Night:Wet +
##     Male:Night:Wet + Level:Male:Night:Wet
##
##                      Df Deviance    AIC
## + Male:Distraction   5   128.04 878.22
## + Level:Distraction 10   123.60 883.78
## + Male:LD            1   145.99 888.17
## <none>                   148.28 888.46
## + Night:LD           1   147.13 889.31
## + Wet:LD             1   148.25 890.43
## + Level:LD           2   146.34 890.52
## + Wet:Distraction    6   146.70 898.88
## + Night:Distraction  6   147.49 899.66
##
## Step:  AIC=878.22
## Count ~ Level + Male + Night + Wet + LD + Distraction + Level:Male +
##     Level:Night + Male:Night + Level:Wet + Male:Wet + Night:Wet +
##     LD:Distraction + Male:Distraction + Level:Male:Night + Level:Male:Wet
+
##     Level:Night:Wet + Male:Night:Wet + Level:Male:Night:Wet
##
##                      Df Deviance    AIC
## + Male:LD            1   125.27 877.44
## <none>                   128.04 878.22
## + Night:LD           1   127.02 879.20
```

```
## + Level:LD           2   125.82 880.00
## + Wet:LD             1   127.99 880.17
## + Level:Distraction  9   118.00 886.18
## + Night:Distraction  6   127.22 889.40
## + Wet:Distraction    6   127.43 889.61
##
## Step:  AIC=877.44
## Count ~ Level + Male + Night + Wet + LD + Distraction + Level:Male +
##     Level:Night + Male:Night + Level:Wet + Male:Wet + Night:Wet +
##     LD:Distraction + Male:Distraction + Male:LD + Level:Male:Night +
##     Level:Male:Wet + Level:Night:Wet + Male:Night:Wet +
Level:Male:Night:Wet
##
##                        Df Deviance    AIC
## + Male:LD:Distraction  5   109.75 871.92
## <none>                     125.27 877.44
## + Level:LD             2   122.36 878.54
## + Night:LD             1   124.48 878.66
## + Wet:LD               1   125.12 879.29
## + Level:Distraction    9   114.95 885.13
## + Night:Distraction    6   124.52 888.70
## + Wet:Distraction      6   124.73 888.91
##
## Step:  AIC=871.92
## Count ~ Level + Male + Night + Wet + LD + Distraction + Level:Male +
##     Level:Night + Male:Night + Level:Wet + Male:Wet + Night:Wet +
##     LD:Distraction + Male:Distraction + Male:LD + Level:Male:Night +
##     Level:Male:Wet + Level:Night:Wet + Male:Night:Wet +
Male:LD:Distraction +
##     Level:Male:Night:Wet
##
##                      Df Deviance    AIC
## <none>                  109.745 871.92
## + Night:LD           1  108.915 873.09
## + Wet:LD             1  109.662 873.84
```

```
## + Level:LD           2  108.174 874.35
## + Level:Distraction  9   98.653 878.83
## + Night:Distraction  6  108.874 883.05
## + Wet:Distraction    6  109.131 883.31
```

```
summary(fit.forward4)
```

```
##
## Call:
## glm(formula = Count ~ Level + Male + Night + Wet + LD + Distraction +
##     Level:Male + Level:Night + Male:Night + Level:Wet + Male:Wet +
##     Night:Wet + LD:Distraction + Male:Distraction + Male:LD +
##     Level:Male:Night + Level:Male:Wet + Level:Night:Wet + Male:Night:Wet +
##     Male:LD:Distraction + Level:Male:Night:Wet, family = poisson,
##     data = Data3)
##
## Deviance Residuals:
##      Min        1Q    Median        3Q       Max
## -2.11264  -0.47249  -0.04127   0.35213   1.98876
##
## Coefficients: (2 not defined because of singularities)
##                                          Estimate Std. Error z value Pr(>|
z|)
## (Intercept)                             7.860e-01  2.517e-01   3.123
0.001793
## Level0                                  2.607e-01  2.215e-01   1.177
0.239331
## Level1                                 -1.595e-01  2.331e-01  -0.684
0.493830
## Malefemale                             -7.522e-01  4.122e-01  -1.825
0.068045
## Nightdaylight                          -1.800e-11  2.182e-01   0.000
1.000000
## Wetdry                                  7.340e-01  1.893e-01   3.876
0.000106
## LDno                                   -2.672e-01  3.553e-01  -0.752
0.451923
## Distraction-3                          -3.929e-01  4.198e-01  -0.936
```

```
0.349302
## Distraction-2                              1.333e+00  2.237e-01   5.959
2.54e-09
## Distraction-1                              8.395e-01  2.340e-01   3.588
0.000334
## Distraction0                               5.478e-01  2.426e-01   2.258
0.023925
## Distraction1                               2.708e-01  2.529e-01   1.071
0.284355
## Distraction2                               7.732e-01  2.468e-01   3.133
0.001729
## Level0:Malefemale                         -3.480e-01  4.180e-01  -0.833
0.405102
## Level1:Malefemale                          5.132e-01  3.948e-01   1.300
0.193559
## Level0:Nightdaylight                       1.214e-01  3.090e-01   0.393
0.694529
## Level1:Nightdaylight                       5.032e-03  3.292e-01   0.015
0.987805
## Malefemale:Nightdaylight                  -1.443e-01  3.995e-01  -0.361
0.717959
## Level0:Wetdry                             -7.340e-01  2.889e-01  -2.540
0.011070
## Level1:Wetdry                             -1.225e-01  2.860e-01  -0.428
0.668367
## Malefemale:Wetdry                         -1.244e-01  3.471e-01  -0.358
0.719982
## Nightdaylight:Wetdry                      -4.082e-02  2.675e-01  -0.153
0.878718
## LDno:Distraction-3                        -1.410e-01  7.093e-01  -0.199
0.842411
## LDno:Distraction-2                        -8.612e-01  3.953e-01  -2.179
0.029368
## LDno:Distraction-1                        -8.791e-01  4.220e-01  -2.083
0.037228
## LDno:Distraction0                         -4.738e-01  4.218e-01  -1.123
0.261247
## LDno:Distraction1                          5.184e-01  4.073e-01   1.273
0.203086
```

```
## LDno:Distraction2                          -4.643e-02  4.144e-01  -0.112
0.910800
## Malefemale:Distraction-3                    1.565e+00  4.938e-01   3.169
0.001528
## Malefemale:Distraction-2                    4.409e-01  3.246e-01   1.358
0.174426
## Malefemale:Distraction-1                   -6.675e-02  3.573e-01  -0.187
0.851798
## Malefemale:Distraction0                     9.891e-02  3.867e-01   0.256
0.798112
## Malefemale:Distraction1                     1.117e+00  3.840e-01   2.908
0.003632
## Malefemale:Distraction2                            NA         NA      NA
NA
## Malefemale:LDno                            -5.886e-01  7.897e-01  -0.745
0.456022
## Level0:Malefemale:Nightdaylight            -3.210e-02  5.849e-01  -0.055
0.956238
## Level1:Malefemale:Nightdaylight            -1.637e-02  5.510e-01  -0.030
0.976301
## Level0:Malefemale:Wetdry                    1.069e-01  5.519e-01   0.194
0.846446
## Level1:Malefemale:Wetdry                   -9.821e-03  4.794e-01  -0.020
0.983656
## Level0:Nightdaylight:Wetdry                -4.964e-02  4.091e-01  -0.121
0.903433
## Level1:Nightdaylight:Wetdry                 3.579e-02  4.033e-01   0.089
0.929280
## Malefemale:Nightdaylight:Wetdry             1.244e-01  4.894e-01   0.254
0.799299
## Malefemale:LDno:Distraction-3              -2.559e-01  1.052e+00  -0.243
0.807900
## Malefemale:LDno:Distraction-2               5.296e-01  8.418e-01   0.629
0.529283
## Malefemale:LDno:Distraction-1               1.131e+00  8.860e-01   1.276
0.201892
## Malefemale:LDno:Distraction0                6.262e-01  9.124e-01   0.686
0.492531
## Malefemale:LDno:Distraction1               -1.152e+00  9.458e-01  -1.218
```

```
0.223158
## Malefemale:LDno:Distraction2                     NA          NA      NA
NA
## Level0:Malefemale:Nightdaylight:Wetdry -8.481e-02  7.771e-01  -0.109
0.913098
## Level1:Malefemale:Nightdaylight:Wetdry -6.339e-03  6.756e-01  -0.009
0.992513
##
## (Intercept)                            **
## Level0
## Level1
## Malefemale                             .
## Nightdaylight
## Wetdry                                 ***
## LDno
## Distraction-3
## Distraction-2                          ***
## Distraction-1                          ***
## Distraction0                           *
## Distraction1
## Distraction2                           **
## Level0:Malefemale
## Level1:Malefemale
## Level0:Nightdaylight
## Level1:Nightdaylight
## Malefemale:Nightdaylight
## Level0:Wetdry                          *
## Level1:Wetdry
## Malefemale:Wetdry
## Nightdaylight:Wetdry
## LDno:Distraction-3
## LDno:Distraction-2                     *
## LDno:Distraction-1                     *
## LDno:Distraction0
## LDno:Distraction1
## LDno:Distraction2
```

```
## Malefemale:Distraction-3                **
## Malefemale:Distraction-2
## Malefemale:Distraction-1
## Malefemale:Distraction0
## Malefemale:Distraction1                 **
## Malefemale:Distraction2
## Malefemale:LDno
## Level0:Malefemale:Nightdaylight
## Level1:Malefemale:Nightdaylight
## Level0:Malefemale:Wetdry
## Level1:Malefemale:Wetdry
## Level0:Nightdaylight:Wetdry
## Level1:Nightdaylight:Wetdry
## Malefemale:Nightdaylight:Wetdry
## Malefemale:LDno:Distraction-3
## Malefemale:LDno:Distraction-2
## Malefemale:LDno:Distraction-1
## Malefemale:LDno:Distraction0
## Malefemale:LDno:Distraction1
## Malefemale:LDno:Distraction2
## Level0:Malefemale:Nightdaylight:Wetdry
## Level1:Malefemale:Nightdaylight:Wetdry
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for poisson family taken to be 1)
##
##     Null deviance: 492.19  on 215  degrees of freedom
## Residual deviance: 109.74  on 168  degrees of freedom
## AIC: 871.92
##
## Number of Fisher Scoring iterations: 4
```

```
fit.forward4$deviance
```

```
## [1] 109.7448
```

Problem 5 Part c

(c) Draw the association graph for the model selected in Part (b). [5 pts]

Problem 5 Part d

(d) Test whether **Distraction** and **LD** are conditionally independent given other variables in the model from Part (b)? [5 pts]

```
fit.noLD <- glm(Count ~. + Level + Male + Night + Wet + LD + Distraction + 
Level*Male +
    Level*Night + Male*Night + Level*Wet + Male*Wet + Night*Wet + 
Male*Distraction + Male*LD + Level*Male*Night +
    Level*Male*Wet + Level*Night*Wet + Male*Night*Wet +
    Level*Male*Night*Wet,
                data = Data3, family = poisson)

summary(fit.noLD)

## 
## Call:
## glm(formula = Count ~ . + Level + Male + Night + Wet + LD + Distraction + 
##     Level * Male + Level * Night + Male * Night + Level * Wet + 
##     Male * Wet + Night * Wet + Male * Distraction + Male * LD + 
##     Level * Male * Night + Level * Male * Wet + Level * Night * 
##     Wet + Male * Night * Wet + Level * Male * Night * Wet, family = 
poisson, 
##     data = Data3)
## 
## Deviance Residuals: 
##     Min       1Q   Median       3Q      Max  
## -2.4056  -0.5835  -0.0731   0.4198   2.4284  
## 
## Coefficients: (1 not defined because of singularities)
##                                            Estimate Std. Error z value Pr(>|
z|)
## (Intercept)                               8.947e-01  2.224e-01   4.024 
5.72e-05
## Level0                                    2.567e-01  2.215e-01   1.159 
0.246314
```

```
## Level1                                     -1.502e-01  2.328e-01  -0.645
0.518872
## Malefemale                                 -8.290e-01  3.941e-01  -2.104
0.035417
## Nightdaylight                               3.682e-16  2.182e-01   0.000
1.000000
## Wetdry                                      7.246e-01  1.891e-01   3.833
0.000127
## LDno                                       -6.326e-01  8.551e-02  -7.397
1.39e-13
## Distraction-3                              -4.131e-01  3.477e-01  -1.188
0.234747
## Distraction-2                               1.086e+00  1.857e-01   5.847
5.01e-09
## Distraction-1                               5.936e-01  1.955e-01   3.037
0.002391
## Distraction0                                4.111e-01  2.001e-01   2.054
0.039990
## Distraction1                                5.480e-01  1.965e-01   2.789
0.005283
## Distraction2                                7.885e-01  1.976e-01   3.991
6.58e-05
## Level0:Malefemale                          -3.862e-01  4.170e-01  -0.926
0.354346
## Level1:Malefemale                           4.550e-01  3.935e-01   1.156
0.247626
## Level0:Nightdaylight                        7.171e-02  3.087e-01   0.232
0.816309
## Level1:Nightdaylight                       -1.019e-02  3.291e-01  -0.031
0.975309
## Malefemale:Nightdaylight                   -1.768e-01  3.991e-01  -0.443
0.657666
## Level0:Wetdry                              -7.246e-01  2.887e-01  -2.510
0.012087
## Level1:Wetdry                              -1.224e-01  2.852e-01  -0.429
0.667739
## Malefemale:Wetdry                          -1.669e-01  3.457e-01  -0.483
0.629201
## Nightdaylight:Wetdry                       -3.144e-02  2.673e-01  -0.118
```

```
0.906386
## Malefemale:Distraction-3                  1.542e+00  4.273e-01   3.608
0.000309
## Malefemale:Distraction-2                  6.633e-01  2.947e-01   2.251
0.024401
## Malefemale:Distraction-1                  2.722e-01  3.211e-01   0.848
0.396578
## Malefemale:Distraction0                   3.265e-01  3.406e-01   0.959
0.337753
## Malefemale:Distraction1                   7.771e-01  3.432e-01   2.264
0.023551
## Malefemale:Distraction2                          NA         NA      NA
NA
## Malefemale:LDno                          -4.381e-01  1.677e-01  -2.612
0.009004
## Level0:Malefemale:Nightdaylight           6.342e-02  5.835e-01   0.109
0.913450
## Level1:Malefemale:Nightdaylight           3.671e-02  5.503e-01   0.067
0.946808
## Level0:Malefemale:Wetdry                  1.971e-01  5.488e-01   0.359
0.719473
## Level1:Malefemale:Wetdry                  6.617e-02  4.769e-01   0.139
0.889659
## Level0:Nightdaylight:Wetdry              -1.345e-02  4.084e-01  -0.033
0.973726
## Level1:Nightdaylight:Wetdry               4.162e-02  4.030e-01   0.103
0.917748
## Malefemale:Nightdaylight:Wetdry           1.669e-01  4.884e-01   0.342
0.732523
## Level0:Malefemale:Nightdaylight:Wetdry -1.999e-01  7.754e-01  -0.258
0.796533
## Level1:Malefemale:Nightdaylight:Wetdry -6.369e-02  6.741e-01  -0.094
0.924728
##
## (Intercept)                            ***
## Level0
## Level1
## Malefemale                             *
```

```
## Nightdaylight
## Wetdry                                      ***
## LDno                                        ***
## Distraction-3
## Distraction-2                               ***
## Distraction-1                               **
## Distraction0                                *
## Distraction1                                **
## Distraction2                                ***
## Level0:Malefemale
## Level1:Malefemale
## Level0:Nightdaylight
## Level1:Nightdaylight
## Malefemale:Nightdaylight
## Level0:Wetdry                               *
## Level1:Wetdry
## Malefemale:Wetdry
## Nightdaylight:Wetdry
## Malefemale:Distraction-3                    ***
## Malefemale:Distraction-2                    *
## Malefemale:Distraction-1
## Malefemale:Distraction0
## Malefemale:Distraction1                     *
## Malefemale:Distraction2
## Malefemale:LDno                             **
## Level0:Malefemale:Nightdaylight
## Level1:Malefemale:Nightdaylight
## Level0:Malefemale:Wetdry
## Level1:Malefemale:Wetdry
## Level0:Nightdaylight:Wetdry
## Level1:Nightdaylight:Wetdry
## Malefemale:Nightdaylight:Wetdry
## Level0:Malefemale:Nightdaylight:Wetdry
## Level1:Malefemale:Nightdaylight:Wetdry
## ---
```

```
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for poisson family taken to be 1)
##
##     Null deviance: 492.19  on 215  degrees of freedom
## Residual deviance: 151.18  on 179  degrees of freedom
## AIC: 891.36
##
## Number of Fisher Scoring iterations: 5
```

```
models <- list(fit.noLD,fit.backward4)
mod.G2 <- sapply(models,function(x)x$deviance)
mod.X2 <- sapply(models,function(x) sum(residuals(x,type="pearson")^2))
mod.df <- sapply(models,function(x) x$df.resid)
mod.pval <- pchisq(mod.G2,mod.df,lower=F)
mod.AIC <- sapply(models,function(x) x$aic)
lackFit <-
data.frame(G2=mod.G2,X2=mod.X2,df=mod.df,pval=mod.pval,AIC=mod.AIC)
rownames(lackFit) <- c("Backward fit without LD int.","Backward fit")
round(lackFit, 10)
```

```
##                                     G2       X2  df      pval      AIC
## Backward fit without LD int. 151.1845 152.7480 179 0.9355263 891.3639
## Backward fit                 179.9114 187.0297 200 0.8429244 878.0908
```

```
res.noLD <- resid(fit.noLD, type = "pearson")/sqrt(1 -lm.influence(fit.noLD)
$hat)
summary(res.noLD)
```

```
##      Min.   1st Qu.    Median      Mean   3rd Qu.      Max.
## -2.155422 -0.604187 -0.081850  0.008573  0.496194  3.057272
```

Problem 5 Part e

(e) Whether the model from Part (b) is equivalent to the main effects logit model or the model with two-way interactions if treating LD as the binary response variable? Justify your answer [5 pts]

The model chosen in part b is (DLdM, LeMNW) which would not be equivalent to either the main effects logit model of the two-way interaction logit model. In order to be equivalent to the main effects logit model, the log linear model would need a five-way interaction term. Similarly, this log linear model is missing terms that would be necessary for the two-way interaction logit model.

```
Tab3orig <- ftable(xtabs(Count~Level+Male+Night+Wet+Distraction+LD,
data=Data3))

Data3.1 <- data.frame(expand.grid(Level=c("Level0","Level1","Level2"),
                                    Male=c("female","male"),
                                    Night=c("daylight","night"),
Wet=c("dry","wet"),

Distraction=c("-3","-2","-1","0","1","2","3"),
                                    LD=c("no","yes")),
                         Count=Tab3orig[1:336])

Tab5.logit <- data.frame(expand.grid(Level=c("Level0","Level1","Level2"),
                                    Male=c("female","male"),
                                    Night=c("daylight","night"),
Wet=c("dry","wet"),

Distraction=c("-3","-2","-1","0","1","2","3")),
                        LDno=Data3.1$Count[1:168],
                         LDyes=Data3.1$Count[169:336])
#Tab5.logit

logit.maineff <- glm(cbind(LDyes,LDno) ~., data=Tab5.logit, family=binomial)
logit.2way <- glm(cbind(LDyes,LDno) ~.^2, data=Tab5.logit, family=binomial)

summary(logit.maineff)

##
## Call:
## glm(formula = cbind(LDyes, LDno) ~ ., family = binomial, data =
Tab5.logit)
##
## Deviance Residuals:
##     Min       1Q   Median       3Q      Max
```

```
## -2.6481  -0.1785   0.0000   0.5966   2.7674
##
## Coefficients:
##                 Estimate Std. Error z value Pr(>|z|)
## (Intercept)    0.892898   0.231316   3.860 0.000113 ***
## LevelLevel0   -0.321813   0.181945  -1.769 0.076937 .
## LevelLevel1   -0.201431   0.181334  -1.111 0.266642
## Malefemale    -0.139255   0.153036  -0.910 0.362848
## Nightdaylight -0.005916   0.149884  -0.039 0.968517
## Wetdry         0.021298   0.150586   0.141 0.887524
## Distraction-3  1.150421   0.435952   2.639 0.008318 **
## Distraction-2  0.435625   0.246136   1.770 0.076751 .
## Distraction-1  0.313688   0.245795   1.276 0.201879
## Distraction0   0.181140   0.240530   0.753 0.451397
## Distraction1  -0.062777   0.228520  -0.275 0.783537
## Distraction2   0.703219   0.280166   2.510 0.012073 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 171.53  on 119  degrees of freedom
## Residual deviance: 149.60  on 108  degrees of freedom
## AIC: 392.83
##
## Number of Fisher Scoring iterations: 4
```

```
summary(logit.2way)
```

```
##
## Call:
## glm(formula = cbind(LDyes, LDno) ~ .^2, family = binomial, data =
Tab5.logit)
##
## Deviance Residuals:
##     Min       1Q   Median       3Q      Max
```

```
## -2.1449  -0.3121   0.0000   0.5206   1.9983
##
## Coefficients:
##                              Estimate Std. Error z value Pr(>|z|)
## (Intercept)                  0.392123   0.534126   0.734   0.4629
## LevelLevel0                  0.052450   0.591218   0.089   0.9293
## LevelLevel1                  0.520415   0.577476   0.901   0.3675
## Malefemale                   0.019303   0.478966   0.040   0.9679
## Nightdaylight               -0.026358   0.468365  -0.056   0.9551
## Wetdry                       0.653732   0.518271   1.261   0.2072
## Distraction-3                2.041041   1.161148   1.758   0.0788 .
## Distraction-2                0.513140   0.719030   0.714   0.4754
## Distraction-1                1.506528   0.686884   2.193   0.0283 *
## Distraction0                 0.131581   0.668342   0.197   0.8439
## Distraction1                 0.723095   0.691899   1.045   0.2960
## Distraction2                 1.630329   0.787706   2.070   0.0385 *
## LevelLevel0:Malefemale       0.005283   0.447148   0.012   0.9906
## LevelLevel1:Malefemale       0.051675   0.418957   0.123   0.9018
## LevelLevel0:Nightdaylight   -0.319774   0.407338  -0.785   0.4324
## LevelLevel1:Nightdaylight    0.043105   0.407072   0.106   0.9157
## LevelLevel0:Wetdry          -0.124919   0.470655  -0.265   0.7907
## LevelLevel1:Wetdry          -0.293017   0.447710  -0.654   0.5128
## LevelLevel0:Distraction-3    1.078188   1.352823   0.797   0.4255
## LevelLevel1:Distraction-3   -0.164251   1.101937  -0.149   0.8815
## LevelLevel0:Distraction-2    0.115183   0.647669   0.178   0.8588
## LevelLevel1:Distraction-2   -0.392775   0.659847  -0.595   0.5517
## LevelLevel0:Distraction-1   -0.119008   0.793071  -0.150   0.8807
## LevelLevel1:Distraction-1   -1.748982   0.687803  -2.543   0.0110 *
## LevelLevel0:Distraction0    -0.265380   0.681514  -0.389   0.6970
## LevelLevel1:Distraction0    -0.620181   0.659636  -0.940   0.3471
## LevelLevel0:Distraction1    -0.210181   0.618881  -0.340   0.7341
## LevelLevel1:Distraction1     0.016475   0.650776   0.025   0.9798
## LevelLevel0:Distraction2    -0.651314   0.875398  -0.744   0.4569
## LevelLevel1:Distraction2    -1.717419   0.878446  -1.955   0.0506 .
## Malefemale:Nightdaylight     0.233208   0.355424   0.656   0.5117
```

```
## Malefemale:Wetdry           -0.586232   0.392738  -1.493   0.1355
## Malefemale:Distraction-3    -0.498872   1.038131  -0.481   0.6308
## Malefemale:Distraction-2    -0.061122   0.590510  -0.104   0.9176
## Malefemale:Distraction-1     0.171256   0.648761   0.264   0.7918
## Malefemale:Distraction0      0.409294   0.533577   0.767   0.4430
## Malefemale:Distraction1     -0.901707   0.542916  -1.661   0.0967 .
## Malefemale:Distraction2      0.352379   0.797187   0.442   0.6585
## Nightdaylight:Wetdry        -0.144597   0.357161  -0.405   0.6856
## Nightdaylight:Distraction-3 -1.013464   1.020749  -0.993   0.3208
## Nightdaylight:Distraction-2  0.371856   0.524871   0.708   0.4787
## Nightdaylight:Distraction-1  0.008069   0.541656   0.015   0.9881
## Nightdaylight:Distraction0  -0.385107   0.566465  -0.680   0.4966
## Nightdaylight:Distraction1   0.026804   0.547265   0.049   0.9609
## Nightdaylight:Distraction2  -0.080559   0.691676  -0.116   0.9073
## Wetdry:Distraction-3        -0.576974   1.081673  -0.533   0.5938
## Wetdry:Distraction-2        -0.425116   0.526189  -0.808   0.4191
## Wetdry:Distraction-1        -1.048577   0.587721  -1.784   0.0744 .
## Wetdry:Distraction0          1.571601   0.649432   2.420   0.0155 *
## Wetdry:Distraction1         -0.434723   0.544392  -0.799   0.4246
## Wetdry:Distraction2         -0.153246   0.800610  -0.191   0.8482
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 171.532  on 119  degrees of freedom
## Residual deviance:  97.446  on  69  degrees of freedom
## AIC: 418.67
##
## Number of Fisher Scoring iterations: 5
```

```
logit.maineff$deviance
```

```
## [1] 149.5997
```

```
logit.2way$deviance
```

```
## [1] 97.44576
```

Problem 5 Part f

(f) Is the model from Part (b) lack of fit? Justify your answer. [5 pts]

The model from part b have a p-value of 0.9999 when testing for fit with the saturated model, so this suggests a good fit. However, when we look at the standardize pearson residuals, the minimum is -inf, suggesting a poor fit. Because of this residual, I would not be able to say confidently that this model has a good fit.

```
models <- list(fit.forward4,fit.4waymod,fit.backward4,fit.main)
mod.G2 <- sapply(models,function(x)x$deviance)
mod.X2 <- sapply(models,function(x) sum(residuals(x,type="pearson")^2))
mod.df <- sapply(models,function(x) x$df.resid)
mod.pval <- pchisq(mod.G2,mod.df,lower=F)
mod.AIC <- sapply(models,function(x) x$aic)
lackFit <-
data.frame(G2=mod.G2,X2=mod.X2,df=mod.df,pval=mod.pval,AIC=mod.AIC)
rownames(lackFit) <- c("Forward fit (b)","4-way mod","Backward fit","Main
effects")
round(lackFit, 5)
```

```
##                        G2       X2  df    pval      AIC
## Forward fit (b) 109.7448 107.5428 168 0.99984 871.9242
## 4-way mod       178.1211 185.1292 185 0.62831 906.3005
## Backward fit    179.9114 187.0297 200 0.84292 878.0908
## Main effects    209.4974 222.4027 203 0.36239 901.6768
```

```
res.forward4 <- resid(fit.forward4, type = "pearson")/sqrt(1-
lm.influence(fit.forward4)$hat)
summary(res.forward4)
```

```
##     Min.  1st Qu.   Median     Mean  3rd Qu.     Max.
##     -Inf -0.53063 -0.05556     -Inf  0.40394  2.56425
```