

Solution: GZLM for Independent Data Practice

November 24, 2003

Problem 1

(a) See output. The LRT can be found in the attached code. The final model is a logistic regression model with gender(newmale) and curriculum(newspc) as predictors. Curriculum(newspc) is kept in the final model although it is non-significant. It is because the purpose of the study is to find out which curriculum is better.

(b) My final model is

$$\log\left(\frac{\pi_i}{1 - \pi_i}\right) = \beta_0 + \beta_{spc}X_{1i} + \beta_{male}X_{2i}.$$

The probability of a collision for a female is the same whatever her GPA is, since there is no GPA in my final model. But it depends on which curriculum she took, since SPC is in my final model. For a female who took SPC, the probability of collision is :

$$\frac{\exp(-1.4109 + 0.1299)}{1 + \exp(-1.4109 + 0.1299)} = 0.217.$$

For a female who took PDLC, the probability of collision is :

$$\frac{\exp(-1.4109)}{1 + \exp(-1.4109)} = 0.196.$$

The probability of a collision for a male is the same whatever his GPA is, since there is no GPA in my final model. But it depends on which curriculum he took, since SPC is in my final model. For a male who took SPC, the probability of collision is :

$$\frac{\exp(-1.4109 + 0.1299 + 0.9259)}{1 + \exp(-1.4109 + 0.1299 + 0.9259)} = 0.412.$$

For a male who took PDLC, the probability of collision is :

$$\frac{\exp(-1.4109)}{1 + \exp(-1.4109 + 0.9259)} = 0.381.$$

(c)

$$\exp(\beta_{male}) = \frac{\text{odds of collision for male}}{\text{odds of collision for female}} = \frac{\pi_m/(1 - \pi_m)}{\pi_f/(1 - \pi_f)}$$

(d) SAS Proc Logistic gave the same fitted regression coefficients and standard errors.

Problem 2

(a) There is no sufficient evidence to support the hypothesized link between SIDS and temperature, since the fitted coefficient of temperature turned out to be nonsignificant with p-value 0.7373.

$$\exp(\beta_{temp}) = \frac{\text{expected death of SIDS at temp} = X + 1}{\text{expected death of SIDS at temp} = X}$$

If our model were plausible, we could have said if the temperature increased by $1^\circ C$, the expected number of SIDS deaths reported would increase to be $e^{0.014} = 1.015 = 101.5\%$ of the original number, which corresponds to an increase of 1.5%.

(b) The parameter estimates changed, but the model didn't fit the data well either. For this model, β_{temp} is interpreted as the increase in expected number of death of SIDS for $1^\circ C$ increase in temperature. Thus $\exp(\beta_{temp})$ is the increase in expected number of death of SIDS in log scale for $1^\circ C$ increase in temperature.

(c) The parameter estimates changed, but the model didn't fit the data well either. For this model, β_{temp} is interpreted as the increase in expected value of log-death of SIDS for $1^\circ C$ increase in temperature.

(d) Poisson regression with log link:

$$\log(EY_i) = \beta_0 + \beta_1 X_{temp}, \text{ where } \overset{\text{we assume}}{Y_i} \sim \text{Poisson.}$$

Poisson regression with identity link:

$$EY_i = \beta_0 + \beta_1 X_{temp}, \text{ where } \overset{\text{we assume}}{Y_i} \sim \text{Poisson.}$$

Simple linear regression with log-transformation:

$$E(\log(Y_i)) = \beta_0 + \beta_1 X_{temp}, \text{ where } \overset{\text{we assume}}{\log(Y_i)} \sim \text{Normal.}$$

```
*****;
*** read in driver education data, create own indicator variables for ***;
*** school, create new categorical variables for spc and gender ***;
*** keep only those observations from year 1 ***;
%include '/home/squid/lynn/sasuser/2col.options.sas';
data dat;
infile '/home/muskie/correlated.data/sids.dat';
input count temp @@;
*****;
*** Poisson regression with log link in genmod ***;
proc genmod data=dat;
model count = temp / dist = poisson link = log;
title "POISSON MODEL with log link";
*****;
*** Poisson regression with identity link in genmod ***;
proc genmod data=dat;
model count = temp / dist = poisson link = identity;
title "POISSON MODEL with identity link";
*****;
*** simple linear regression of log transformed counts ***;
data dat;
set dat;
Incount = log(count+1/6);
*****;
proc glm data=dat;
model Incount = temp;
title "NORMAL MODEL on log transformed count";
```

logist.practice.sas

```

*****
*** read in driver education data, create own indicator variables for ***
*** school, create new categorical variables for SPC and gender ***
*** keep only those observations from year 1 ***
*****
%include '/home/squid/lynn/sasuser/2col.options.sas';
data dat;
  infile '/home/muskie/correlated.data/driver.dat' firstobs=2;
  input id collide year spc school male gpa;
  if school = 1 then do;
    sch1 = 1; sch2 = 0; end;
  else if school = 2 then do;
    sch1 = 0; sch2 = 1; end;
  else if school = 3 then do;
    sch1 = 0; sch2 = 0; end;
  if spc = 1 then newspc = 1;
  else if spc = 0 then newspc = 2;
  if male = 1 then newmale = 1;
  else if male = 0 then newmale = 2;

  if year = 1;
/*
*** fit full model in genmod
*****
proc genmod data=dat;
  class newspc school newmale;
  model collide = newspc|school|newmale|gpa / dist = bin link = logit;
  title "FULL MODEL, all interactions";
*****
*** fit reduced model in genmod
*****
proc genmod data=dat;
  class newspc school newmale;
  model collide = newspc school newmale gpa / dist = bin link = logit;
  title "MAIN EFFECTS ONLY MODEL";
  *** LRT = -2(-180.5993 - (-174.9419)) = 11.31
  *** df = 23 - 5 = 18
  *** p = 0.88
  /
***
*** fit final model in genmod
*****
proc genmod data=dat;
  class newmale;
  model collide = newmale gpa / dist = bin link = logit;
  title "FINAL MODEL";
  *** LRT = -2(-181.8684 - (-174.9419)) = 13.85
  *** df = 23 - 2 = 21
  *** p = 0.86
  /
*** fit final model in logistic
***
proc logistic data=dat descending;
  model collide = male gpa;
  title "FINAL MODEL";
  /
  *** FROM GENMOD ***

```

Analysis of Parameter Estimates

Parameter	DF	Estimate	Std Err	ChiSquare	Pr>Chi
-----------	----	----------	---------	-----------	--------

```

*** FROM LOGISTIC ***
Analysis of Maximum Likelihood Estimates

```

Variable	DF	Parameter Estimate	Standard Error	Wald Chi-Square	Pr > Chi-Square	Standardized Estimate	Odds Ratio
INTERCEPT	1	-2.0324	0.6826	8.8643	0.0029	8.8643	0.0029
NEWMALE	1	0.9159	0.2553	12.8669	0.0003	12.8669	0.0003
NEWMALE	2	0	0	0.0000	0.0000	0	0.0000
GPA	1	0.3142	0.2496	1.5847	0.2081	1.5847	0.2081
SCALE	0	1.0000	0	0.0000	0.0000	1.0000	0.0000

logist.practice.1st

21:39 Thursday, November 20, 2003 5

21:39 Thursday, November 20, 2003 6

FINAL MODEL
The GENMOD Procedure

Model Information

WORK DAT
Distribution Binomial
Link Function Logit
Dependent Variable collide
Observations Used 300
Probability Modeled Pr(collide = 1)

Class Level Information

Class Levels Values
newmale 2 1 2

Response Profile

Ordered Level	Ordered Value	Count
1	0	202
2	1	98

Criteria For Assessing Goodness Of Fit

Criterion	DF	Value	Value/DF
Deviance	297	365.0678	1.2292
Scaled Deviance	297	365.0678	1.2292
Pearson Chi-Square	297	299.8479	1.0096
Scaled Pearson X2	297	299.8479	1.0096
Log Likelihood		-182.5339	

Algorithm converged.

Analysis Of Parameter Estimates

Parameter	DF	Estimate	Standard Error	Wald 95% Confidence Limits	Chi-Square	Pr > ChiSq
Intercept	1	-1.4109	0.4278	-2.2494 -0.5724	10.88	0.0010
newspc	1	0.1299	0.2523	-0.3646 0.6244	0.27	0.6067
newmale	1	0.9259	0.2547	0.4266 1.4251	13.21	0.0003
scale	2	0	0.0000	0.0000 1.0000		

NOTE: The scale parameter was held fixed.

MAIN EFFECTS ONLY MODEL
The GENMOD Procedure

Model Information

WORK DAT
Distribution Binomial
Link Function Logit
Dependent Variable collide
Observations Used 300
Probability Modeled Pr(collide = 1)

Class Level Information

Class Levels Values
newspc 2 1 2
school 3 1 2 3
newmale 2 1 2

Response Profile

Ordered Level	Ordered Value	Count
1	0	202
2	1	98

Criteria For Assessing Goodness Of Fit

Criterion	DF	Value	Value/DF
Deviance	294	361.1986	1.2286
Scaled Deviance	294	361.1986	1.2286
Pearson Chi-Square	294	301.0020	1.0238
Scaled Pearson X2	294	301.0020	1.0238
Log Likelihood		-180.5993	

Algorithm converged.

Analysis Of Parameter Estimates

Parameter	DF	Estimate	Standard Error	Wald 95% Confidence Limits	Chi-Square	Pr > ChiSq
Intercept	1	-1.8095	0.6974	-3.1765 -0.4426	6.73	0.0095
newspc	1	-0.1396	0.2545	-0.6384 0.3592	0.30	0.5833
newspc	2	0	0.0000	0.0000 0.0000		
school	1	-0.4280	0.3131	-1.0416 0.1856	1.87	0.1716
school	2	-0.3573	0.3070	-0.9590 0.2444	1.35	0.2444
school	3	0	0.0000	0.0000 0.0000		
newmale	1	0.9089	0.2576	0.4041 1.4137	12.45	0.0004
newmale	2	0	0.0000	0.0000 0.0000		
gpa	1	0.3541	0.2519	-0.1397 0.8478	1.98	0.1598
Scale	0	1.0000	0.0000	1.0000 1.0000		

NOTE: The scale parameter was held fixed.



The LOGISTIC Procedure

Model Information

Data Set WORK.DAT
 Response Variable collide
 Number of Response Levels 2
 Number of Observations 300
 Link Function Logit
 Optimization Technique Fisher's scoring

Response Profile

Ordered Value	collide	Total Frequency
1	1	98
2	0	202

Model Convergence Status

Convergence criterion (GCONV=1E-8) satisfied.

Model Fit Statistics

Criterion	Intercept Only	Intercept and Covariates
AIC	381.076	371.068
SC	384.779	382.179
-2 Log L	379.076	365.068

Testing Global Null Hypothesis: BETA=0

Test	Chi-Square	DF	Pr > ChiSq
Likelihood Ratio	14.0079	2	0.0009
Score	13.8614	2	0.0010
Wald	13.5047	2	0.0012

Analysis of Maximum Likelihood Estimates

Parameter	DF	Estimate	Standard Error	Chi-Square	Pr > ChiSq
Intercept	1	-1.1511	0.2282	25.4402	<.0001
male	1	0.9359	0.2547	13.2117	0.0003
spc	1	-0.1299	0.2523	0.2651	0.6067

The LOGISTIC Procedure

Odds Ratio Estimates

Effect	Point Estimate	95% Wald Confidence Limits
male	2.524	1.532 4.158
spc	0.878	0.536 1.440

Association of Predicted Probabilities and Observed Responses

	Percent Concordant	Somers' D	Percent Discordant	Gamma	Percent Tied	Tau-a
Pairs	50.2	0.243	26.0	0.318	23.8	0.107
C	19796	0.621				

poisson practice1st

POISSON MODEL with log link

20:24 Thursday, November 20, 2003 1

POISSON MODEL with identity link

20:24 Thursday, November 20, 2003 2

The GENMOD Procedure

Model Information

Data Set	WORK.DAT
Distribution	Poisson
Link Function	Log
Dependent Variable	count
Observations Used	242

Criteria For Assessing Goodness Of Fit

Criterion	DF	Value	Value/DF
Deviance	240	314.2785	1.3095
Scaled Deviance	240	314.2785	1.3095
Pearson Chi-Square	240	292.4263	1.2184
Scaled Pearson X2	240	292.4263	1.2184
Log Likelihood		-241.5885	

Algorithm converged.

Analysis Of Parameter Estimates

Parameter	DF	Estimate	Standard Error	Wald 95% Confidence Limits	Chi-Square	Pr > ChiSq
Intercept	1	-0.3625	0.9193	-2.1643 1.4392	0.16	0.6933
temp	1	0.0147	0.0437	-0.0711 0.1004	0.11	0.7373
Scale	0	1.0000	0.0000	1.0000 1.0000		

NOTE: The scale parameter was held fixed.

The GENMOD Procedure

Model Information

Data Set	WORK.DAT
Distribution	Poisson
Link Function	Identity
Dependent Variable	count
Observations Used	242

Criteria For Assessing Goodness Of Fit

Criterion	DF	Value	Value/DF
Deviance	240	314.3032	1.3096
Scaled Deviance	240	314.3032	1.3096
Pearson Chi-Square	240	292.3141	1.2180
Scaled Pearson X2	240	292.3141	1.2180
Log Likelihood		-241.6009	

Algorithm converged.

Analysis Of Parameter Estimates

Parameter	DF	Estimate	Standard Error	Wald 95% Confidence Limits	Chi-Square	Pr > ChiSq
Intercept	1	0.7180	0.7797	-0.8101 2.2462	0.85	0.3571
temp	1	0.0109	0.0372	-0.0620 0.0838	0.09	0.7694
Scale	0	1.0000	0.0000	1.0000 1.0000		

NOTE: The scale parameter was held fixed.

poisson.practice.lst

NORMAL MODEL on log transformed count

20:24 Thursday, November 20, 2003 3

NORMAL MODEL on log transformed count

20:24 Thursday, November 20, 2003 4

The GLM Procedure
Number of Observations 242

The GLM Procedure
Dependent Variable: Incount

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	1	0.4351605	0.4351605	0.31	0.5767
Error	240	334.2555001	1.3927213		
Corrected Total	241	334.6906606			

R-Square	Coef Var	Root MSE	Incount Mean
0.001300	-246.4122	1.180140	-0.478929

Source	DF	Type I SS	Mean Square	F Value	Pr > F
temp	1	0.43516048	0.43516048	0.31	0.5767

Source	DF	Type III SS	Mean Square	F Value	Pr > F
temp	1	0.43516048	0.43516048	0.31	0.5767

Parameter	Estimate	Standard Error	t Value	Pr > t
Intercept	-1.072377117	1.06438146	-1.01	0.3147
temp	0.028353704	0.05072461	0.56	0.5767