

Name _____

Laboratory #10
The Generalized Linear Model - Logistic Regression (Erin Carruthers, 2009)

The purpose of this lab is to develop the capacity to analyze binomial data in a model based format. The Generalized Linear Model lets you apply all of the structural models you have learned (regression, single factor ANOVA, multiple regression, multiple factor ANOVA, ANCOVA) to binomial data. In this lab you will calculate observed proportions, odds, and odds ratios. Then you will use parameter estimates to calculate expected values of proportions, odds, and odds ratios.

Once you have completed the lab write-up, you should have

- ability to place binomial data in a structure suitable for model based analysis
- ability to use a statistical package to execute any model you write
- ability to relate proportions to odds, and to relate odds to odds ratios
- ability to interpret results of model based analysis of binomial data

The data for this lab comes from a classic paper on analysis of binomial data.

Blood Pressure bP	Sample Size N	Observed Disease Nhd
111.5	156	3
121.5	252	17
131.5	284	12
141.5	271	16
151.5	139	12
161.5	85	8
176.5	99	16
191.5	43	8

BP = blood pressure in units of pressure (mm of mercury)

N = number of men age 40-59 in the Framingham longitudinal study

Nhd = number of men developing heart disease during Framingham cohort study

Data by Cornfield, J. 1962. Federated Proceedings 21: 58-61
as reported by

Agresti A. 1990 Categorical Data Analysis pp 93-95

This lab is readily completed in a spreadsheet or in MathCad.

Parameter estimates can be obtained with any statistical package.

Name _____

Section 1. Preliminary calculations - observed proportions, odds, and odds ratios.

Using the numbers in bold (below), calculate morbidity (proportion with heart disease).

Then calculate Odds and Odds Ratios from morbidity.

BP = 111.5 is the reference category, and hence OR = 1 by definition.

Several calculations are shown, to check yours.

Fill in the 3 blank cells (bP = 191.5) in the table below.

Table 1

Blood Pressure	Sample Size	Observed Disease	Morbidity =Nhd/N	Odds =p/(1-p)	Odds Ratio
bP	N	Nhd	Morb	Odds	OR
111.5	156	3	0.019	0.020	1.000
121.5	252	17	0.067	0.072	3.689
131.5	284	12			
141.5	271	16			
151.5	139	12			4.819
161.5	85	8			
176.5	99	16			
191.5	43	8			

Section 2. Model definition.

Now, we write the model for estimating the change in odds of heart disease with increase in blood pressure.

$$Odds = e^{\eta} + \text{binomial error}$$

$$\eta = \beta_{ref} + \beta_{bP} bP$$

Section 3. Model execution.

Next we use a logistic regression routine to estimate the parameters β_{ref} and β_{bP} . This can be done with a logistic regression routine, or with a generalized linear model routine with binomial error and log link.

Statistical packages differ as to how the response variable is defined.

In some packages (e.g. SAS) the response variable is Nhd/N

In some packages (e.g. Minitab) the two components of the response variable (Nhd and N) are entered separately.

Name _____

Here is code for SAS, R, and Minitab. The minitab code can be generated using the pull-down menus.

```
Proc Genmod;  
  Model Nhd/N = bP/  
  Link=logit dist=binomial type1 type3;
```

SAS command file

```
HdModel <- glm(Nhd/N ~ bP, family = binomial(link = logit),  
weights = N, data = Cornfield)  
summary(HdModel)  
plot(HdModel)
```

R command lines

```
MTB > BLogistic 'Nhd' 'N' = 'bP';                               as of 2002  
SUBC>   ST;  
SUBC>   Logit;  
SUBC>   Brief 2.
```

Minitab command lines

```
Click Stat  
  Click Regression  
    Click Binary Logistic Regression  
      Click Success, place column of Nhd,  
      Click trials, place column of N  
      Click Model, place column with bP  
      Click Storage (optional)  
      Click Pearson residuals, Event probability, ok
```

Minitab sequence to generate command lines

The estimates from your package should be close to those below, estimated from R/SPlus.

Parameter estimates:

$$\beta_{ref} = -6.107$$
$$\beta_{bP} = 0.0245$$

Calculate the odds of heart disease at bP = 111.5 mmHg $e^{(-6.107)} =$ _____

Calculate the rate of change in odds of heart disease per unit change in blood pressure

$$e^{(0.0245)} = \text{_____ \% / mmHg}$$

Name _____

Section 4. More calculations - Expected odds ratios, expected odds, and expected proportions.

Parameter estimates:
 $\beta_{ref} = -6.107$
 $\beta_{bP} = 0.0245$

Expected Odds Ratio, from parameter
 Using the bP numbers in bold, calculate ΔbP
 Using ΔbP and β_{bP} , calculate the expected Odds Ratio, E(OR), from the formula
 Several calculations are shown, to check yours.
 Fill in the 2 blank cells (bP = 191.5) in Table 2.

Table 2

Blood Pressure	ΔbP	Expected Odds Ratio $E(OR) = e^{\beta_{bP} * \Delta bP}$
111.5	0	1.000
121.5	10	1.278
131.5	20	
141.5		
151.5		
161.5		
176.5	65	4.916
191.5		

Expected Odds, from parameters.
 Using the bP numbers in bold (below), and both parameter estimates, calculate the expected odds at each observed blood pressure level. Several calculations are shown, to check yours.
 Fill in the 4 blank cells (bP = 191.5) in the table below.

Table 3
 $Odds = e^{\eta} + \text{binomial error}$ $\eta = \beta_{ref} + \beta_{bP} * bP$

bP	b_{ref}	$B_{bP} * bP$	η	$e^{\eta} = E(Odds)$
111.5	-6.107	2.730	-3.377	0.034
121.5	-6.107	2.975	-3.132	0.044
131.5				
141.5				
151.5				0.091
161.5				
176.5				
191.5				

Expected Odds, from parameters.
 Using the parameter estimates, you can calculate the expected odds for any blood pressure.
 Calculate the expected (predicted) odds at an unobserved value, bP = 181.5

E(OR) = _____ E(Odds) = _____

Expected morbidity, from parameters

The expected morbidity can be computed from the expected odds, which were computed from the model parameters.

The formula for morbidity, from odds, is

$$p = \text{odds}/(1+\text{odds}).$$

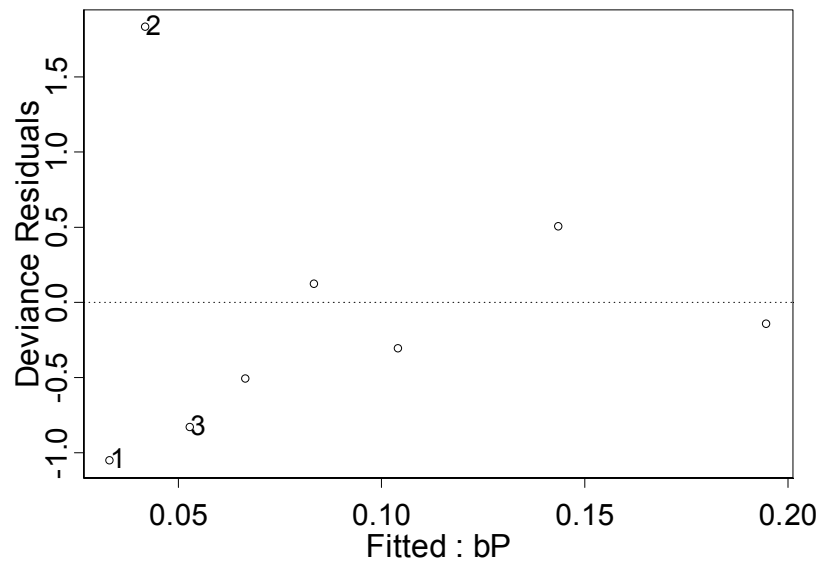
Several calculations are shown, to check yours. Fill in the 2 blank cells (bP = 191.5) in Table 4.

Table 4

bP	Obs(Morb)	E(Morb)
111.5	0.019	0.033
121.5	0.067	0.042
131.5	0.042	0.053
141.5		
151.5		
161.5		
176.5		
191.5		

Section 5. Model evaluation.

Model evaluation is similar to that for the general linear model (normal error and identity link) in that we are concerned primarily with homogeneity of residuals. And we are also concerned with normality of the residuals. To evaluate these assumptions we use scaled residuals, not raw residuals (raw residual = Observed - Fitted Value).



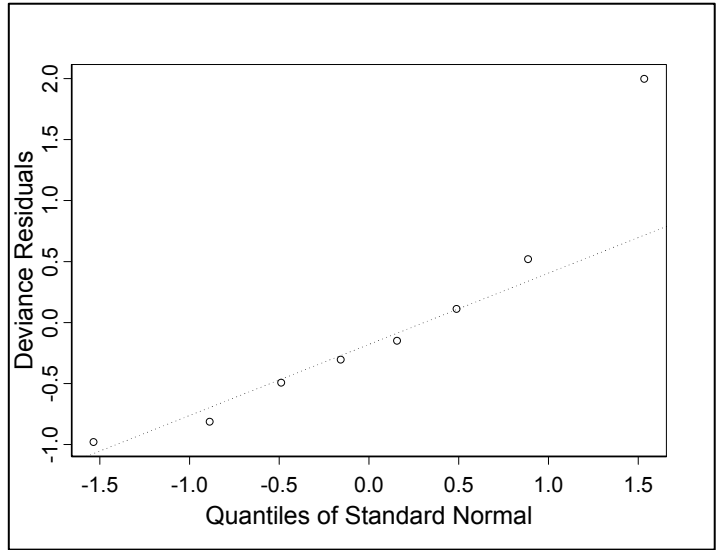
Here is the plot of deviance residuals (a suitably scaled residual) against the fitted values (fitted proportions).

The first, second, and third observations deviate from the model, due primarily to the second observation.

Section 5. Model evaluation.

When we look at the normality plot, we see again that there is an outlier (the second observation, morbidity at bP = 121.5).

Comparison of the observed to expected morbidity shows that morbidity at bP = 121.5 is unusually high, compared to the model expectation.



Set up a new table with this line removed.

Recompute the parameter estimates and display them here:

$$\beta_{ref} = \underline{\hspace{2cm}} \quad \beta_{bP} = \underline{\hspace{2cm}}$$

Using the new parameter estimates, calculate the expected odds and odds ratios at bP = 191.5

$$E(OR) = \underline{\hspace{2cm}} \quad E(Odds) = \underline{\hspace{2cm}}$$

Before leaving the lab today, make sure you know how to copy a spreadsheet and place it into a printable document.

Write-up for this lab.

1. Fill in the blank cells in Tables 1, 2, 3, and 4.
Fill in the blanks on page 2-6.
2. Report the parameter estimates before and after removal of $bP = 121.5$. Comment on the direction of change in the parameter estimates.
Report the expected morbidity, expected odds, and expected odds ratios at $bP = 191.5$, before and after removal of the value at $bP = 121.5$. Comment on the direction of change in the parameter estimates.
3. Finally, generalize and state the direction of change in the expected morbidity, expected odds, and expected odds ratio at any level of bP , when the regression parameter β_{bP} increases or decreases. Be sure to check your generalization against several values of β_{bP} .

Extra: The formula for a proportion, from odds, is

$$p = \text{odds}/(1+\text{odds}).$$

Using algebra, obtain this formula from the formula for odds:

$$\text{odds} = p/(1-p)$$