APPLIED STATISTICS

Logistic Regression for Binomial Counts

Dr Tao Zou

Research School of Finance, Actuarial Studies & Statistics
The Australian National University

Last Updated: Mon Oct 9 21:12:17 2017

Overview

- Motivating Example

- Logistic Regression for Binomial Counts

- Model Diagnostics

  1. Logit proportion verus explanatory variable plot.

  2. Pearson residual plot.

  3. Deviance goodness-of-fit test.

References

1. **F.L. Ramsey and D.W. Schafer** (2012)
   Chapter 21 of *The Statistical Sleuth*

2. ANU STAT3015 Lecture Notes

3. The slides are made by **R Markdown**.
   http://rmarkdown.rstudio.com

Example: Anaesthetic Data (Con'd)

(Taken from STAT3015 notes.)

Thirty patients were given a particular anaesthetic at various predetermined concentrations for 15 minutes before a stimulus was applied.

The response variable was simply an indication as to whether the patient responded to the stimulus in any way.

"Response" is 1 if the patient responded to the stimulus.

Standard Data Form

```
setwd('~/Desktop/Research/AppliedStat2017/L12')
a=read.csv('anaesthetic.csv');a
```

```
##    Concentration Response
## 1            0.8        1
## 2            0.8        1
## 3            0.8        1
## 4            0.8        1
## 5            0.8        1
## 6            0.8        1
## 7            0.8        0
## 8            1.0        1
## 9            1.0        1
## 10           1.0        1
## 11           1.0        1
## 12           1.0        0
## 13           1.2        1
## 14           1.2        1
## 15           1.2        0
## 16           1.2        0
## 17           1.2        0
## 18           1.2        0
## 19           1.4        1
## 20           1.4        1
## 21           1.4        0
## 22           1.4        0
## 23           1.4        0
## 24           1.4        0
## 25           1.6        0
## 26           1.6        0
## 27           1.6        0
## 28           1.6        0
## 29           2.5        0
## 30           2.5        0
```

Some Terminologies to Interpret Standard Data

"Trial": Thirty patients correspond to thirty trials for this anaesthetic experiment.

Different trials are independent.

"Success": one trial is called "success" if in this trial, the patient responds to the stimulus, or equivalently, Response $= 1$.

"Failure": one trial is called "failure" if in this trial, the patient does not respond to the stimulus, or equivalently, Response $= 0$.

Probability of "Success": the probability that one trial is a "success", **or equivalently**, the probability that the patient responds to the stimulus in the trial, **or equivalently**, the probability that Response $= 1$.

Aggregated Data Form

Note that in the data, we have repeated concentration (X) values. Can we aggregate response Y by X?

The number of "successes" (i.e. Response $= 1$ here) given some $X = x$:

```
attach(a)
tapply(Response, Concentration,sum)
```

```
## 0.8   1 1.2 1.4 1.6 2.5
##   6   4   2   2   0   0
```

The number of "trials" (equivalent to patients here) given some $X = x$:

```
tapply(Response, Concentration,length)
```

```
## 0.8   1 1.2 1.4 1.6 2.5
##   7   5   6   6   4   2
```

The aggregated dataset is

```
b=data.frame(Concentration=unique(Concentration),BinomialCount=tapply(Response, Concentration,sum),
             TotalNumber=tapply(Response, Concentration,length));b
```

```
##     Concentration BinomialCount TotalNumber
## 0.8           0.8             6           7
## 1             1.0             4           5
## 1.2           1.2             2           6
## 1.4           1.4             2           6
## 1.6           1.6             0           4
## 2.5           2.5             0           2
```

Binomial Count

In a lot of real data, we only have **the aggregated data form** instead of the standard data form. Specifically in this example, we have

1. at a specific concentration, we have **the number of "successes"** (the number of patients that respond to the stimulus), which is called **binomial count**, denoted by Z;

2. at a specific concentration, we have **the total number of "trials"** (the total number of patients given this specific level of anaesthetic concentration), denoted by M.

Logistic regression for binomial counts deals with the particular case of the aggregated data, and is used to model the binomial count response Z.

However, the total number of "trials" M should also be observed such that we can employ this model.

Some Terminologies to Interpret Aggregated Data

A typical row for an aggregated dataset:

Explanatory	Binomial Count	Total Number
X	Z	M

Z is the number of "successes" given some specific X.

M is the number of "trials" given some specific X.

Given X, the probability of one "success" in a trial is denoted by $\pi \in [0, 1]$.

The binomial count Z is modelled by Binomial(M, π) distribution.
Specifically,

$$\mathrm{P}(Z = z) = \binom{M}{z} \pi^z (1-\pi)^{M-z}, \binom{M}{z} = \frac{M(M-1)\cdots(M-z+1)}{z(z-1)\cdots 1},$$

for $z = 0, \cdots, M$.

Given X, the proportion of "successes" is Z/M, which can be used to estimate π when M is large.

Overview of This Course

	Continuous X + Categorical X
Continuous Y	MLR + Indicator Variables
Two-Category Y	Binary Logistic Regression + Indicator Variables
Multicategory Y - Nominal	Nominal Response Regression + Indicator Variables
Multicategory Y - Ordinal	Ordinal Response Regression + Indicator Variables
Binomial Count Z	**Binomial Logistic Regression + Indicator Variables**

Binomial Logistic Regression Model Assumptions

1. **Binomial distribution**: There is a Binomial distributed (sub)population of responses Z for given values of the explanatory variables $(X_1 = x_1, \cdots, X_k = x_k)$. That means if we let $X = (X_1, \cdots, X_k)$, the probability that $Z = z$ given X is

$$\mathrm{P}(Z = z) = \binom{M}{z} \pi^z (1 - \pi)^{M-z}.$$

2. **Generalised Linearity**: The transformation of the probability of "success" π falls on a linear function of the explanatory variables

$$g(\pi) = \beta_0 + \beta_1 X_1 + \cdots + \beta_k X_k, \text{ for } X = (X_1, \cdots, X_k),$$

where $g(u) = \log\{u/(1 - u)\}$, which is called logit link function.

Binary Logistic Regression Model Assumptions (Con'd)

Remark: the inverse function of the logit link function is

$$g^{-1}(v) = \frac{e^v}{1 + e^v} \in [0, 1].$$

Then

$$\pi = g^{-1}(\beta_0 + \beta_1 X_1 + \cdots + \beta_k X_k) \in [0, 1],$$

which is consistent with the range $\pi \in [0, 1]$.

3. **Independence**: Observations

$$(X_{1,1}, \cdots X_{k,1}, Z_1, M_1),$$

$$\vdots$$

$$(X_{1,m}, \cdots X_{k,m}, Z_m, M_m),$$

are independent, where m is the sample size.

The interpretation of the binomial logistic regression model stays largely the same as for the binary logistic regression model.

Estimation, z-Test and CI

The likelihood function for the observations and the MLE can be obtained.
The inferential tools for generalised linear model can be used.

```
attach(b)
Y=cbind(BinomialCount,TotalNumber-BinomialCount)
fit=glm(Y~Concentration,family=binomial(link=logit))
summary(fit)
```

```
##
## Call:
## glm(formula = Y ~ Concentration, family = binomial(link = logit))
##
## Deviance Residuals:
##       0.8        1       1.2       1.4       1.6       2.5
## -0.20147  0.45367 -0.56890  0.70000 -0.81838 -0.04826
##
## Coefficients:
##               Estimate Std. Error z value Pr(>|z|)
## (Intercept)      6.469      2.419   2.675  0.00748 **
## Concentration   -5.567      2.044  -2.724  0.00645 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 15.4334  on 5  degrees of freedom
## Residual deviance:  1.7321  on 4  degrees of freedom
## AIC: 13.811
##
## Number of Fisher Scoring iterations: 5
```

Drop-in-Deviance χ^2-Test

$$H_0 : \beta_1 = 0 \leftrightarrow H_a : \text{ otherwise.}$$

```
#drop in deviance test
#reduced model
fitr=glm(Y~1,family=binomial(link=logit))
anova(fitr,fit,test='Chisq')
```

```
## Analysis of Deviance Table
##
## Model 1: Y ~ 1
## Model 2: Y ~ Concentration
##    Resid. Df Resid. Dev Df Deviance  Pr(>Chi)
## 1          5    15.4334
## 2          4     1.7321  1   13.701 0.0002143 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Fitted Probabilities of "Success"

Using MLE $\hat{\beta}_0, \cdots, \hat{\beta}_k$, the estimated (fitted) probability of "success" is given by:

$$
\begin{aligned}
\hat{\pi} &= g^{-1}(\hat{\beta}_0 + \hat{\beta}_1 X_1 + \cdots + \hat{\beta}_k X_k) \\
&= \frac{e^{\hat{\beta}_0 + \hat{\beta}_1 X_1 + \cdots + \hat{\beta}_k X_k}}{1 + e^{\hat{\beta}_0 + \hat{\beta}_1 X_1 + \cdots + \hat{\beta}_k X_k}}.
\end{aligned}
$$

It is worth noting that

$$
g(\hat{\pi}) = \log\left(\frac{\hat{\pi}}{1 - \hat{\pi}}\right) = \hat{\beta}_0 + \hat{\beta}_1 X_1 + \cdots + \hat{\beta}_k X_k,
$$

where function g is the logit link function. Hence the logit transformation of fitted probabilities should be on a straight line.

When we talk about "fitted", X is usually from the training dataset (see Lecture Notes 8).

When X_{new} is from the new dataset or the test dataset, we actually talk about prediction.
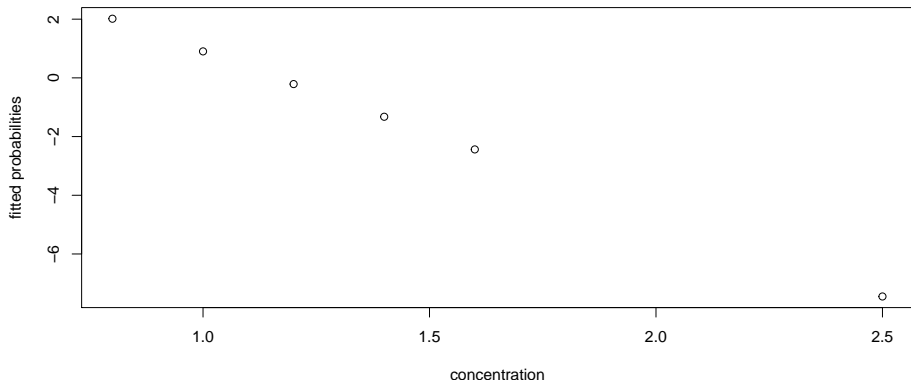
Predicted Probabilities of "Success"

The predicted probability of "success" is given by

$$
\begin{aligned}
\hat{\pi} &= g^{-1}(\hat{\beta}_0 + \hat{\beta}_1 X_{1,\text{new}} + \cdots + \hat{\beta}_k X_{k,\text{new}}) \\
&= \frac{e^{\hat{\beta}_0 + \hat{\beta}_1 X_{1,\text{new}} + \cdots + \hat{\beta}_k X_{k,\text{new}}}}{1 + e^{\hat{\beta}_0 + \hat{\beta}_1 X_{1,\text{new}} + \cdots + \hat{\beta}_k X_{k,\text{new}}}}.
\end{aligned}
$$

Example: Anaesthetic Data (Con'd)

```
plot(Concentration,log(fit$fitted.values/(1-fit$fitted.values)),
     xlab="concentration",ylab="fitted probabilities")
```


The logit transformations of fitted probabilities are on a straight line.

Example: Anaesthetic Data (Con'd)

By using this example, we might be interested in predicting the probability of "success" if an anaesthetic at a new concentration of 1.5 is given. The forecast of probability is:

```
Xnew=data.frame(Concentration=1.5)
predict(fit,Xnew,type='response')


##         1
## 0.1322204
```

Fitted Values of Response and Prediction of Response

A typical row for an aggregated data:

Explanatory	Binomial Count	Total Number
X	Z	M

The binomial count Z is modelled by Binomial(M, π) distribution. Specifically,

$$\mathrm{P}(Z = z) = \binom{M}{z} \pi^z (1 - \pi)^{M-z}, \text{ for } z = 0, \cdots, M.$$

Based on the properties of the Binomial distribution, the mean of response Z is given by

$$\mu\{Z|X\} = M\pi.$$

Using MLE $\hat{\beta}_0, \cdots, \hat{\beta}_k$, the fitted values of response Z are given by:

$$\hat{Z} = \hat{\mu}\{Z|X\} = M\hat{\pi}.$$

Example: Anaesthetic Data (Con'd)

By using this example, we might be interested in predicting the number of "successes" if an anaesthetic at a new concentration of 1.5 is given. Suppose the number of "trials" is 5 in this case. The forecast should be

```
Mnew=5
Mnew*predict(fit,Xnew,type='response')
```

```
##         1
## 0.6611018
```

However, it is not an integer. So we consider

```
round(Mnew*predict(fit,Xnew,type='response'),0)
```

```
## 1
## 1
```

roughly as the prediction of response Z.

1. Logit Proportion verus Explanatory Variable Plot

Recall that in the Binomial logistic regression model assumptions, the **logit** transformation of the probability of "success" π falls on a linear function of the explanatory variables

$$g(\pi) = \beta_0 + \beta_1 X_1 + \cdots + \beta_k X_k, \text{ for } X = (X_1, \cdots, X_k),$$

where $g(u) = \log\{u/(1-u)\}$ is the **logit** link function. Here the probability of "success" π depends on unknown parameters and hence is also unknown.

For a typical row for an aggregated data:

Explanatory	Binomial Count	Total Number
X	Z	M

The proportion of "successes" is Z/M given Z, which can be used to estimate π when M is large.

Hence it can be useful to plot the logit proportion $g(Z/M)$ versus explanatory variable (e.g., X_1).

1. Logit Proportion verus Explanatory Variable Plot (Con'd)

The plot should show a straight line. Otherwise the model assumption is violated. See Lecture Notes 3 for solutions to this problem.

Note: $g(Z/M)$ is undefined for Z/M equal to 0 or 1. We need to add (substract) a small quantity to values of 0 (1).

```
proportion=BinomialCount/TotalNumber
proportion=ifelse(proportion==0, 0.01, proportion)
proportion=ifelse(proportion==1, 1-0.01, proportion)
logitprop=log(proportion/(1-proportion))
plot(Concentration,logitprop)
```
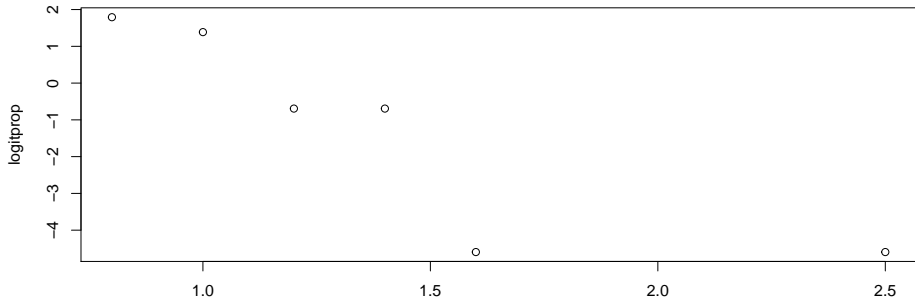
Review: Studentized Residuals for Continuous Response Y

Recall residual: $\text{res}_i = \hat{\mathcal{E}}_i = Y_i - \hat{Y}_i$ for observation i, where $\hat{Y}_i$ is the fitted value of response.

One can obtain

$$\text{SD}(\text{res}_i) = \sigma\sqrt{1 - h_i}, \text{ and}$$

$$\text{SE}(\text{res}_i) = \hat{\sigma}\sqrt{1 - h_i}, \text{ where } \hat{\sigma} = \sqrt{\frac{\sum_{i=1}^{n}\text{res}_i^2}{n - k - 1}}.$$

A studentized residual is a residual divided by its standard error, namely

$$\text{studres}_i = \frac{\text{res}_i}{\text{SE}(\text{res}_i)}$$

Using studentized residuals allows the residuals to be viewed on the same scale.

These studentized residuals are roughly standard normally distributed $[N(0, 1)]$, if the observation is from the MLR model with the all the assumptions satisfied.

Pearson Residuals for Binomial Count Z

The residual for Binomial count Z is defined as: $\text{res}_i = Z_i - \hat{\hat{Z}}_i$ for observation i, where $\hat{\hat{Z}}_i$ is the fitted value of response.

One can obtain

$$\text{SD}(\text{res}_i) = \sqrt{M_i \pi_i (1 - \pi_i)}, \text{ and}$$

$$\text{SE}(\text{res}_i) = \sqrt{M_i \hat{\pi}_i (1 - \hat{\pi}_i)}, \text{ where } \hat{\pi}_i = \frac{e^{\hat{\beta}_0 + \hat{\beta}_1 X_{1,i} + \cdots + \hat{\beta}_k X_{k,i}}}{1 + e^{\hat{\beta}_0 + \hat{\beta}_1 X_{1,i} + \cdots + \hat{\beta}_k X_{k,i}}}.$$

A Pearson residual is a residual divided by its standard error, namely

$$\text{Peares}_i = \frac{\text{res}_i}{\text{SE}(\text{res}_i)}$$

Using Pearson residuals allows the residuals to be viewed on the same scale.

For large values of M_i (>5), these Pearson residuals are roughly standard normally distributed [$N(0, 1)$], if the observation is from the Binomial logistic model with the all the assumptions satisfied.

2. Pearson Residual Plot

Due to the nature of the $N(0,1)$ distribution, most values of the $N(0,1)$ distribution concentrate in the middle around 0.

Hence, if Peares_i falls into the two tails of the $N(0,1)$ distribution, namely $|\mathrm{Peares}_i|$ is too large, then it is unlikely that observation i is from the Binomial logistic model with the all the assumptions satisfied.

> $\mathrm{Peares}_i > 1.96$ (97.5 % quantile of $N(0,1)$), or
> $\mathrm{Peares}_i <\text{-} 1.96$ (2.5 % quantile of $N(0,1)$).
> $\Rightarrow$
> Peares_i falls into the two tails of the $N(0,1)$ distribution.
> $\Rightarrow$
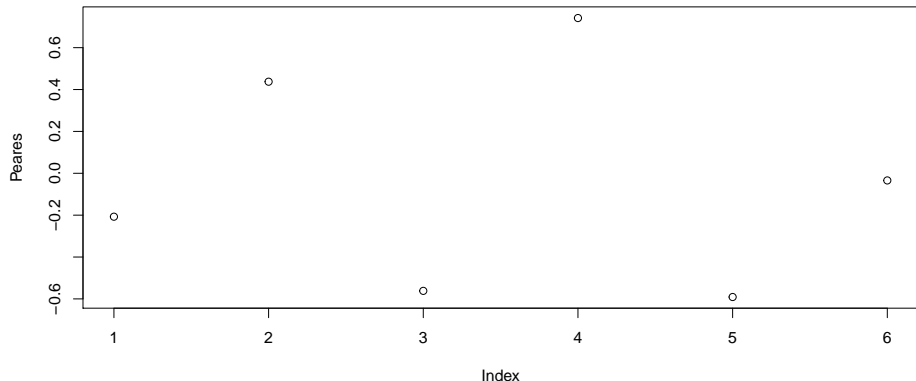> $|\mathrm{Peares}_i|$ is too large.
> $\Rightarrow$
> Observation i is **an outlier**.

Sometimes we use 2 instead of 1.96 for simplicity.

2. Pearson Residual Plot (Con'd)

```
Peares=residuals(fit,type="pearson")
plot(Peares)
abline(h=1.96,col='red')
abline(h=-1.96,col='red')
```


The Pearson residuals do not indicate any problems - none are outside the range of $[-1.96, 1.96]$ and no pattern is evident.

3. Deviance Goodness-of-Fit Test

H_0 : Binomial model $g(\pi) = \beta_0 + \beta_1 X_1 + \cdots + \beta_k X_k$ is appropriate $\leftrightarrow$

H_a : The Binomial logistic model is not appropriate.

The test statistic is

$$TS = \text{deviance of model } g(\pi) = \beta_0 + \beta_1 X_1 + \cdots + \beta_k X_k.$$

This test statistic should be compared to approximately χ^2_{m-k-1} distribution, where k is the number of explanatory variables.

The p-value is

$$p\text{-value} = P(S > TS), \text{ where } S \sim \chi^2_{m-k-1}.$$

If p-value $< \alpha \Rightarrow$ reject H_0; p-value $\geq \alpha \Rightarrow$ not reject H_0.

3. Deviance Goodness-of-Fit Test (Con'd)

The output from the summary() output can be used to perform the goodness-of-fit test.

```
summaryfit=(summary(fit))
summaryfit
```

```
##
## Call:
## glm(formula = Y ~ Concentration, family = binomial(link = logit))
##
## Deviance Residuals:
##       0.8        1       1.2      1.4      1.6      2.5
## -0.20147  0.45367  -0.56890  0.70000  -0.81838  -0.04826
##
## Coefficients:
##                Estimate Std. Error z value Pr(>|z|)
## (Intercept)       6.469      2.419   2.675  0.00748 **
## Concentration    -5.567      2.044  -2.724  0.00645 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 15.4334  on 5  degrees of freedom
## Residual deviance:  1.7321  on 4  degrees of freedom
## AIC: 13.811
##
## Number of Fisher Scoring iterations: 5
```

3. Deviance Goodness-of-Fit Test (Con'd)

The test statistic and the numer of degress of freedom are

```
summaryfit$deviance
```

```
## [1] 1.732136
```

```
summaryfit$df.residual
```

```
## [1] 4
```

The *p*-value for the test is given by

```
1 - pchisq(summaryfit$deviance, summaryfit$df.residual)
```

```
## [1] 0.7848721
```

```
detach(b)
```

We cannot reject the null. There is no evidence that the model is inappropriate.