APPLIED STATISTICS

Multicategory Response Regression

Dr Tao Zou

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

Last Updated: Mon Oct 2 15:18:23 2017

Overview

- Multicategory Response Variables

- Nominal Response Regression Models

- Ordinal Response Regression Models

References

1. **H. Wang** (2008)
   Chapter 5 of *Applied Business Statistical Analysis*
2. **C.R. Bilder & T.M. Loughin** (2015)
   Chapter 3 of *Analysis of Categorical Data with R*
3. The slides are made by **R Markdown**.
   http://rmarkdown.rstudio.com

Multicategory Response Variables

Example 1: Y takes values of "red" and "yellow".
Example 2: Y takes values of "red", "yellow" and "blue".
Example 3: Y takes values of "disagree", "neutral" and "agree".

- As for Example 1, we can define an indicator variable I_Y such that I_Y is 1 if $Y =$"red"; otherwise 0. Binay logistic regression models can be used to model the response I_Y.

- The response Y in Example 2 can be called nominal response. Example 1 is a special case of nominal responses with only two categories.

- The response Y in Example 3 can be called ordinal response.

What is the difference between Example 2 and Example 3?

- In Example 3, we can have an order for the categories "disagree"<"neutral"<"agree". Hence, sometimes in a questionnaire, we can set 1="disagree", 2="neutral" and 3="agree". Note that

$$\text{"neutral"} - \text{"disagree"} \neq \text{"agree"} - \text{"neutral"}.$$

  This means there is no numerical meaning for "disagree", "neutral" and "agree". However, it is not the case for the count data, e.g., the binomial count or the Poisson count introduced later in this course.

- In Example 2, we do not have an order for the categories "red", "yellow" and "blue". It does not matter to set 1="red", 2="yellow" and 3="blue", or 2="red", 1="yellow" and 3="blue".

The Difference between Example 2 and Example 3

- The model to predict the response in Example 2 is called nominal response regression model.
- The model to predict the response in Example 3 is called ordinal response regresssion model.
- In fact, the ordinal response is a special case of the nominal response. Hence the nominal response regression models can also be used for the ordinal response.
- However, the ordinal response has more information. If we use the nominal response regression models for the ordinal response, definitely we lose some model accuracy.

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**

Nominal Response Regression Models

Binary Model Assump

Review that in the binary logistic regression models,

odds

$$\frac{P(Y=1|X)}{P(Y=0|X)} = \frac{\pi(X)}{1-\pi(X)} = \frac{\mathrm{P}(Y=1|X)}{1-\mathrm{P}(Y=1|X)} = e^{\beta_0+\beta_1 X_1+\cdots+\beta_k X_k},$$

where $\pi(X) = \mathrm{P}(Y=1|X)$ is the probability that $Y=1$ given X. In the following lectures, we simplify the notation by $\pi = \mathrm{P}(Y=1)$ for convenience.

$|X$

For multicategory response, denote the response categories $c = 1, \cdots, C$ and $\pi_c = \mathrm{P}(Y=c)$ be the probability that category c happens.

$\pi_1 + \cdots + \pi_C = 1$

Suppose $c=1$ to be the baseline level category. Then the nominal response regression model (baseline-category logit model) is

$\frac{\pi_2}{\pi_1}, \frac{\pi_3}{\pi_1}, \frac{\pi_4}{\pi_1} \cdots \frac{\pi_C}{\pi_1}$

In total $(k+1)(C-1)$

$$\frac{\pi_c}{\pi_1} = e^{\beta_{c0}+\beta_{c1}X_1+\cdots+\beta_{ck}X_k}, \textbf{ only } \text{for } c = 2, \cdots, C.$$

$\boxed{c=2}$ $\frac{\pi_2}{\pi_1} = e^{\beta_{20}+\beta_{21}X_1+\cdots+\beta_{2k}X_k}$ $(k+1)$

$\frac{\pi_3}{\pi_1}$

The likelihood function for the observations and the MLE can be obtained.

Example: Wheat Kernels Data

The presence of sprouted or diseased kernels in wheat can reduce the value of a wheat producer's entire crop.

Example: Wheat Kernels Data (Con'd)

It is important to identify these kernels after being harvested but prior to sale.

To facilitate this identification process, automated systems have been developed to separate healthy kernels from the rest.

Improving these systems requires better understanding of the measurable ways in which healthy kernels differ from kernels that have sprouted prematurely or are infected with a fungus ("Scab").

To this end, Martin et al. (1998) conducted a study examining numerous physical properties of kernels — density, hardness, size, weight, and moisture content — measured on a sample of wheat kernels from two different classes of wheat, hard red winter (hrw) and soft red winter (srw).

predict

Response

Each kernel's condition was also classified as "Healthy," "Sprout," or "Scab" by human visual inspection.

R Code

```
rm(list=ls())
setwd('~/Desktop/Research/AppliedStat2017/L11')
wheat=read.csv('wheat.csv')
head(wheat)
```

```
##   class  density hardness    size  weight moisture    type
## 1   hrw 1.349253 60.32952 2.30274 24.6480 12.01538 Healthy
## 2   hrw 1.287440 56.08972 2.72573 33.2985 12.17396 Healthy
## 3   hrw 1.233985 43.98743 2.51246 31.7580 11.87949 Healthy
## 4   hrw 1.336534 53.81704 2.27164 32.7060 12.11407 Healthy
## 5   hrw 1.259040 44.39327 2.35478 26.0700 12.06487 Healthy
## 6   hrw 1.300258 48.12066 2.49132 33.2985 12.18577 Healthy
```

```
levels(wheat$type)
```

baseline level by default

1. wheat $ type is factor

2.

```
## [1] "Healthy" "Scab"    "Sprout"
```

1 2 3

Estimation and CI

categorical X

```
#install.packages('nnet')
library(nnet)
mod.fit<-multinom(formula=type~class+density+hardness+size+
                  weight+moisture,data=wheat)
```

```
## # weights:  24 (14 variable)
## initial  value 302.118379
## iter  10 value 234.991271
## iter  20 value 192.127549
## final  value 192.112352
## converged
```

```
summary(mod.fit)
```

$\pi_1 = P(Y=1)$

$\pi_2 = P(Y=2)$

$\pi_3 = P(Y=3)$

$\pi_1+\pi_2+\pi_3=1$

1, 2, 3

"Healthy"

Y X_1 X_2 X_3 X_4 X_5 X_6

```
## Call:
## multinom(formula = type ~ class + density + hardness + size +
##     weight + moisture, data = wheat)
##
## Coefficients:
##        (Intercept)   classsrw   density    hardness      size     weight
## Scab      30.54650 -0.6481277 -21.59715 -0.01590741 1.0691139 -0.2896482
## Sprout    19.16857 -0.2247384 -15.11667 -0.02102047 0.8756135 -0.0473169
##           moisture
## Scab    0.10956505
## Sprout -0.04299695
##
## Std. Errors:
##        (Intercept)  classsrw  density    hardness      size     weight
## Scab      4.289865 0.6630948 3.116174 0.010274587 0.7722862 0.06170252
## Sprout    3.767214 0.5009199 2.764306 0.008105748 0.5409317 0.03697493
##         moisture
## Scab   0.1548407
## Sprout 0.1127188
##
## Residual Deviance: 384.2247
## AIC: 412.2247
```

Scab (2), Sprout (3)

$$\frac{\pi_2}{\pi_1} = e^{\boxed{\beta_{20} + \boxed{\beta_{21}} X_1 + \cdots + \beta_{26} X_6}} \to 7$$

$$\frac{\pi_3}{\pi_1} = e^{\beta_{30} + \boxed{\beta_{31}} X_1 + \cdots + \beta_{36} X_6} \to 7$$

14

H_0: X_1 is not needed

$\Leftrightarrow \beta_{21} = \beta_{31} = 0$

$$\text{classsrw} = \begin{cases} 1, & \text{if class} = \text{srw} \\ 0, & \text{if class} = \text{hrw} \end{cases}$$

Indicator

use "hrw" as the baseline

/26

Drop-in-Deviance χ^2-Test

$$\mathrm{H}_0 : X_j \text{ is not needed} \leftrightarrow \mathrm{H}_1 : X_j \text{ is } \cancel{\text{not}} \text{ needed}$$

in the model that already contains other variables, is equivalent to

$$\mathrm{H}_0 : \beta_{2j} = \cdots = \beta_{Cj} = 0 \leftrightarrow \mathrm{H}_1 : \text{ at least one of } \beta_{2j}, \cdots, \beta_{Cj} \text{ is not zero.}$$

correspond to x_j

```r
library(car)
Anova(mod.fit)
```

```
## Analysis of Deviance Table (Type II tests)
##
## Response: type
##          LR Chisq Df Pr(>Chisq)
## class       0.964  2     0.6175
## density    90.555  2  < 2.2e-16 ***
## hardness    7.074  2     0.0291 *
## size        3.211  2     0.2008
## weight     28.230  2  7.411e-07 ***
## moisture    1.193  2     0.5506
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

X_1

Prediction

```r
xnew=data.frame(class='hrw',density=1.4,hardness=50,size=2.5,
                weight=30,moisture=12)
predict(mod.fit,newdata=xnew,type='probs')
```

```
##     Healthy        Scab     Sprout
## 0.937317507 0.005239616 0.057442877
```

```r
predict(mod.fit,newdata=xnew,type='class')
```

```
## [1] Healthy
## Levels: Healthy Scab Sprout
```

Ordinal Response Regression Models

Suppose the response categories $c = 1, \cdots, C$ and category 1 < category 2 < $\cdots$ < category C.

order

$$\mathrm{P}(Y \le c) = \overset{Y=1}{\pi_1} + \cdots + \overset{Y=C}{\pi_c} \text{ for } c = 1, \cdots, C.$$

The ordinal response regression model is

odds that $Y \le C$

$$\frac{\mathrm{P}(Y \le c)}{1 - \mathrm{P}(Y \le c)} = \frac{\pi_1 + \cdots + \pi_c}{\pi_{c+1} + \cdots + \pi_C} = e^{\beta_{c0} + \overset{\beta_{c1}}{\beta_1} X_1 + \cdots + \overset{\beta_{ck}}{\beta_k} X_k},$$

only for $c = 1, \cdots, C - 1$. Note that

$Y \le 1 \to \beta_{10}$

$Y \le 2 \to \beta_{20}$

...

$$\boxed{\mathrm{P}(Y \le C) \equiv 1.}$$

The likelihood function for the observations and the MLE can be obtained.

Example: Wheat Kernels Data (Con'd)

$$\text{scab } (Y = 1) < \text{sprout } (Y = 2) < \text{healthy } (Y = 3)$$

```
levels(wheat$type)
```

```
## [1] "Healthy" "Scab"    "Sprout"
```

Correct

```
wheat$type.order<-factor(wheat$type
                         ,levels=c('Scab','Sprout','Healthy'))
levels(wheat$type.order)
```

```
## [1] "Scab"    "Sprout"  "Healthy"
```

1 2 3

Estimation and CI

```r
library(MASS)
mod.fit.ord<-polr(formula=type.order~class+density+hardness+size+
                  weight+moisture,data=wheat,method="logistic")
summary(mod.fit.ord)
```

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


## Call:
## polr(formula = type.order ~ class + density + hardness + size +
##     weight + moisture, data = wheat, method = "logistic")
##
## Coefficients:
##              Value Std. Error t value
## classsrw   0.17370   0.391764  0.4434
## density   13.50534   1.713009  7.8840
## hardness   0.01039   0.005932  1.7522
## size      -0.29253   0.413095 -0.7081
## weight     0.12721   0.029996  4.2411
## moisture  -0.03902   0.088396 -0.4414
##
## Intercepts:
##                 Value   Std. Error t value
## Scab|Sprout     17.5724  2.2460     7.8237
## Sprout|Healthy  20.0444  2.3395     8.5677
##
## Residual Deviance: 422.4178
## AIC: 438.4178
```

$\hat{\beta}_j$ $\quad SE(\hat{\beta}_j)$ $\quad$ z-value

$\hat{\beta}_{10}$ $\quad \hat{\beta}_{20}$

$H_0: \beta_j = 0$

$TS = \dfrac{\hat{\beta}_j}{SE(\hat{\beta}_j)}$

$\beta_1 \to X_1$

$\beta_2 \to X_2$

$\vdots$

$\beta_6 \to X_6$

$\to$ 6 slopes

$\dfrac{P(Y\le 1)}{P(Y\ge 2)}$

$$\frac{P(Y\le 1)}{1-P(Y\le 1)} = e^{\beta_{10} + \beta_1 X_1 + \cdots \beta_6 X_6}$$

$$\frac{P(Y\le 2)}{1-P(Y\le 2)} = e^{\beta_{20} + \beta_1 X_1 + \cdots \beta_6 X_6}$$

Drop-in-Deviance χ^2-Test

For $j = 1, \cdots, k$,

$\mathrm{H}_0 : X_j$ is not needed $\leftrightarrow$ $\mathrm{H}_1 : X_j$ is ~~not~~ needed

in the model that already contains other variables, is equivalent to

$$\mathrm{H}_0 : \beta_j = 0 \leftrightarrow \mathrm{H}_1 : \beta_j \neq 0$$

z - test
χ^2- test
based on page 18

```
Anova(mod.fit.ord)
```

```
## Analysis of Deviance Table (Type II tests)
##
## Response: type.order
##           LR Chisq Df Pr(>Chisq)
## class        0.197  1    0.65749
## density     98.437  1  < 2.2e-16 ***
## hardness     3.084  1    0.07908 .
## size         0.499  1    0.47982
## weight      18.965  1  1.332e-05 ***
## moisture     0.195  1    0.65872
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

X_1

Prediction

```
xnew=data.frame(class='hrw',density=1.4,hardness=50,size=2.5,
                weight=30,moisture=12)
predict(mod.fit.ord,newdata=xnew,type='probs')
```

```
##       Scab    Sprout    Healthy
## 0.01129879 0.10793873 0.88076248
```

```
predict(mod.fit.ord,newdata=xnew,type='class')
```

```
## [1] Healthy
## Levels: Scab Sprout Healthy
```