Logistic Regression

Chun Li, PhD

Division of Biostatistics

Department of Population and Public Health Sciences

University of Southern California

June 12-16, 2023

Basic Concepts for Binary Outcomes

A **binary outcome** is an outcome variable with two outcome categories.

- Examples: Disease vs. non-disease, success vs. failure, yes vs. no, etc.

We designate one of the two outcome categories as the **event of interest**.

- We often code the event of interest as 1 and the other category as 0.

Suppose p is the **probability** for the event of interest with $0 < p < 1$. Then

- The **odds** of the event is $\frac{p}{1-p}$, and the **log-odds** is $\log(\frac{p}{1-p})$.
- The function $\log(\frac{p}{1-p})$ is called the **logit** function and is often expressed as $\text{logit}(p)$.
- Here, $\log()$ refers to the natural logarithm.

Odds Ratios

The probability of the event of interest can differ under different conditions.

The **odds-ratio** (OR) for Condition 1 versus Condition 2 is $OR_{12} = \frac{p_1/(1-p_1)}{p_2/(1-p_2)}$, where p_k be the probability of the event of interest under Condition k.

- The OR for Condition 2 versus Condition 1 is $OR_{21} = \frac{p_2/(1-p_2)}{p_1/(1-p_1)} = 1/OR_{12}$.

- An OR is always <u>between two conditions</u>, and it depends on the order of the conditions.

- If the two conditions are defined by a binary variable (e.g., sex), with its coding well specified (e.g., 0=F, 1=M), one may say "the OR for sex".

- If the coding is unclear, "the OR for sex" is ambiguous because it could be the OR for man vs. woman or the OR for woman vs. man; if the former is 2 then the latter is 0.5.

- If the conditions are defined by a variable with more than two values (e.g., age), a phrase like "the OR for age" may not make sense (unless under a specific context, say, when linear effect is assumed and the scale of age is specified).

Logistic Regression

Logistic regression is to model a binary outcome against a set of predictors $\boldsymbol{X} = (X_1, \ldots, X_p)$.

Specifically, we model the outcome probability (not the outcome itself) as a function of $\boldsymbol{X}$:

$$p = \sigma(\beta_0 + \beta^T \boldsymbol{X}) = \frac{e^{\beta_0 + \beta^T \boldsymbol{X}}}{1 + e^{\beta_0 + \beta^T \boldsymbol{X}}},$$

or equivalently,

$$\text{logit}(p) = \log\left(\frac{p}{1-p}\right) = \beta_0 + \beta^T \boldsymbol{X}.$$

Here $\beta = (\beta_1, \ldots, \beta_p)$ is a vector of coefficients, and $\beta^T \boldsymbol{X} = \beta_1 X_1 + \cdots + \beta_p X_p$.

$\sigma(t) = \frac{e^t}{1+e^t}$ is the inverse of the logit function: If $p = \sigma(t)$ then $t = \text{logit}(p)$.

Maximum Likelihood Estimation

Suppose we have *independent* observations with

- Y, the outcome (1=disease, 0=no disease)
- X_1, age (in years)
- X_2, smoking status (1=smoker, 0=non-smoker)

disease	age	smoking
1	25	1
0	30	1
0	27	0
1	25	0
$\vdots$	$\vdots$	$\vdots$

The likelihoods for the individuals are

$$P(Y = 1 | X_1 = 25, X_2 = 1) = \frac{e^{\beta_0 + 25\beta_1 + \beta_2}}{1 + e^{\beta_0 + 25\beta_1 + \beta_2}}$$

$$P(Y = 0 | X_1 = 30, X_2 = 1) = \frac{1}{1 + e^{\beta_0 + 30\beta_1 + \beta_2}}$$

$$P(Y = 0 | X_1 = 27, X_2 = 0) = \frac{1}{1 + e^{\beta_0 + 27\beta_1}}$$

$$P(Y = 1 | X_1 = 25, X_2 = 0) = \frac{e^{\beta_0 + 25\beta_1}}{1 + e^{\beta_0 + 25\beta_1}}$$

$$\cdots$$

The **overall likelihood** is the product of these individual likelihoods. It is then maximized to obtain the **MLEs** for $\beta_0, \beta_1, \beta_2$. (An iterative algorithm will be used.)

Example R-Code and Results

Suppose the full dataset is ($n = 12$):

disease	age	smoking
1	25	1
0	30	1
0	27	0
1	25	0
0	35	1
0	22	1
1	38	0
1	40	0
1	32	1
0	30	1
0	29	0
1	37	0

```
mod1 = glm(disease ~ age + smoking, data=dd, family=binomial)
summary(mod1)$coef
```

```
##               Estimate Std. Error    z value  Pr(>|z|)
## (Intercept) -3.3391574  4.1972874 -0.7955513 0.4262929
## age          0.1262221  0.1312073  0.9620052 0.3360470
## smoking     -1.0657874  1.2922082 -0.8247799 0.4094965
```

Note that $e^{-1.0658} = 0.344$ is the estimated OR for smoking *when age is fixed* (i.e., within every age group). It is often different from the marginal OR for smoking.

```
table(dd$disease, dd$smoking)   ## marginal distribution
```

```
##
##     0 1
##   0 2 4
##   1 4 2
```

The marginal OR for smoking is $\frac{2 \times 2}{4 \times 4} = 0.25$, which can also be obtained below:

```
mod2 = glm(disease ~ smoking, data=dd, family=binomial)
exp(summary(mod2)$coef[2,1])
```

```
## [1] 0.25
```

Unadjusted vs. Adjusted ORs

If we fit a logistic regression model of Y on a single variable X_1, as in

$$\text{logit}(P(Y = 1 | X_1)) = \beta_0 + \beta_1 X_1,$$

then e^{β_1} is the **unadjusted** OR (or marginal OR) for X_1.

If we fit a logistic regression model of Y on two or more variables $X_1, \cdots, X_p$, as in

$$\text{logit}(P(Y = 1 | X_1, \ldots, X_p)) = \beta_0 + \beta_1 X + \cdots + \beta_p X_p,$$

then e^{β_1} is the **adjusted** OR for X_1 after adjusting for the effects of the other variables (or we can say e^{β_1} is the **adjusted** OR for X_1 while fixing the values of the other variables).

Interpretation of Model Coefficients

Suppose our model is

$$\text{logit}(p) = \log\left(\frac{p}{1-p}\right) = \log(\text{odds}) = \beta_0 + \beta_1 X_1 + \beta_2 X_2.$$

where X_1 is age (in years) and X_2 is smoking status (1=smoker, 0=non-smoker).

Then (elaboration on next slide)

- β_1 is the <u>log-OR</u>, and e^{β_1} is the OR, for X_1 (technically, e^{β_1} is the OR for one unit increment of X_1 while keeping X_2 value fixed).

- β_2 is the <u>log-OR</u>, and e^{β_2} is the OR, for X_2 (while keeping X_1 value fixed).

- β_0 is the <u>log-odds</u>, and e^{β_0} is the odds, for the "baseline subject" (which corresponds to $X_1 = 0$ and $X_2 = 0$). In this example, the "baseline subject" does not exist in reality.

Interpretation of Model Coefficients (cont'd)

Consider a subject with $X_1 = x_1$ and $X_2 = x_2$, and a second subject with $X_1 = x_1 + 1$ and $X_2 = x_2$. The OR for the second subject versus the first one is

$$\frac{\text{odds for second subject}}{\text{odds for first subject}} = \frac{e^{\beta_0 + \beta_1(x_1 + 1) + \beta_2 x_2}}{e^{\beta_0 + \beta_1 x_1 + \beta_2 x_2}} = e^{\beta_1}.$$

Thus, β_1 is the log-OR for the second subject versus the first one, or β_1 is the *log-OR for increasing X_1 by one unit while keeping all other variables fixed*. It depends on the scale of X_1.

Consider a "baseline subject" with $X_1 = 0$ and $X_2 = 0$. The odds for the subject is

$$e^{\beta_0 + \beta_1 \cdot 0 + \beta_2 \cdot 0} = e^{\beta_0}.$$

Thus, β_0 is the log-odds for the "baseline subject" (which may not exist in reality).

Interpretation of Model Coefficients (cont'd)

Consider a model with two binary predictors: sex (0=F, 1=M), smoking status (0=NS, 1=S),

$$\log\left(\frac{p}{1-p}\right) = \beta_0 + \beta_1 \cdot \text{sex} + \beta_2 \cdot \text{smoking}$$

$$= \begin{cases} \beta_0, & \text{female non-smoker;} \\ \beta_0 + \beta_1, & \text{male non-smoker;} \\ \beta_0 \qquad + \beta_2, & \text{female smoker;} \\ \beta_0 + \beta_1 + \beta_2, & \text{male smoker.} \end{cases}$$

- e^{β_0} is the <u>odds</u> for the "baseline" group of female non-smokers (who have the "baseline" sex and "baseline" smoking status *according to our coding*).

- e^{β_1} is the <u>OR</u> for sex (male vs. female). The OR for sex is the **same** among non-smokers and among smokers.

- e^{β_2} is the <u>OR</u> for smoking (smoking vs. non-smoking). The OR for smoking is the **same** among females and among males.

Suppose We Have X_3 (sex)

Suppose we also have X_3 (sex, 1=man, 0=woman):

disease	age	smoking	sex
1	25	1	1
0	30	1	0
0	27	0	1
1	25	0	0
0	35	1	0
0	22	1	1
1	38	0	1
1	40	0	0
1	32	1	1
0	30	1	1
0	29	0	0
1	37	0	1

```
mod3 = glm(disease ~ smoking + sex, data=dd, family=binomial)
summary(mod3)$coef
```

```
##              Estimate Std. Error   z value  Pr(>|z|)
## (Intercept)  0.1811935  1.053815  0.1719405 0.8634843
## smoking     -1.6824533  1.357387 -1.2394796 0.2151679
## sex          1.1326672  1.377308  0.8223776 0.4108620
```

Here, $e^{-1.6825} = 0.186$ is the estimated OR for smoking *when sex is fixed*, and $e^{1.1327} = 3.10$ is the estimated OR for sex *when smoking is fixed*.

```
mod4 = glm(disease ~ age + smoking + sex, data=dd, family=binomial)
summary(mod4)$coef
```

```
##               Estimate Std. Error    z value  Pr(>|z|)
## (Intercept) -4.5030382  4.5615830 -0.9871657 0.3235614
## age          0.1449225  0.1386784  1.0450260 0.2960109
## smoking     -1.3479594  1.4248541 -0.9460333 0.3441316
## sex          1.3523022  1.4547410  0.9295828 0.3525872
```

Here, $e^{1.3523} = 3.87$ is the estimated OR for sex *when both age and smoking are fixed*.

Interaction Terms in Logistic Regression

Now consider **an interaction model** between sex and smoking,

$$\log\left(\frac{p}{1-p}\right) = \beta_0 + \beta_1 \cdot \text{sex} + \beta_2 \cdot \text{smoking} + \beta_3 \cdot \text{sex} \cdot \text{smoking}$$

$$= \begin{cases} \beta_0, & \text{female non-smoker;} \\ \beta_0 + \beta_1, & \text{male non-smoker;} \\ \beta_0 \quad\;\; + \beta_2, & \text{female smoker;} \\ \beta_0 + \beta_1 + \beta_2 + \beta_3, & \text{male smoker.} \end{cases}$$

- e^{β_0} is the <u>odds</u> for the "baseline" group of female non-smokers.
- e^{β_1} is the <u>OR</u> for sex among non-smokers. (The OR for sex among smokers is $e^{\beta_1+\beta_3}$.)
- e^{β_2} is the <u>OR</u> for smoking among females. (The OR for smoking among males is $e^{\beta_2+\beta_3}$.)
- e^{β_3} is a <u>ratio of ORs</u>:

$$e^{\beta_3} = \frac{\text{OR for sex among smokers}}{\text{OR for sex among non-smokers}} = \frac{\text{OR for smoking among males}}{\text{OR for smoking among females}}.$$

Interaction Terms in Logistic Regression (cont'd)

When there is an interaction term between X_1 and X_2,

- β_1 and β_2 do not reflect the overall effects of X_1 and X_2, That is, they are not "main effects".

- β_0, β_1, and β_3 are not comparable because they are different concepts and are on different scales:
    - β_0 is a log-odds.
    - β_1 and β_2 are log-ORs.
    - β_3 is the difference of two log-ORs.

- It is meaningless to make statements like $\beta_1 = \beta_3$ even though they may happen to have the same value.

- It is meaningless to state that "the interaction effect is stronger than the main effect".

R-code for Interaction Models

```
mod3 = glm(disease ~ smoking * sex, data=dd,
           family=binomial)
summary(mod3)$coef
```

```
##                  Estimate  Std. Error      z value   Pr(>|z|)
## (Intercept)  6.931472e-01    1.224745  5.659523e-01 0.5714262
## smoking     -1.925922e+01 4612.202148 -4.175709e-03 0.9966683
## sex         -1.659121e-15    1.732051 -9.578940e-16 1.0000000
## smoking:sex  1.856607e+01 4612.202419  4.025424e-03 0.9967882
```

We can check the coefficient estimates
manually.

```
## disease dist. in baseline of female non-smokers:
## odds = 2/1 = 2 = exp(6.931e-1)
with(dd, table(disease[smoking==0 & sex==0]))
```

```
##
## 0 1
## 1 2
```

```
## disease-smoking distribution among females:
## OR = (0/2)/(2/1) = 0 = exp(-1.926e1)
with(dd, table(disease[sex==0], smoking[sex==0]))
```

```
##
##     0 1
## 0 1 2
## 1 2 0
```

```
## disease-sex distribution among non-smokers:
## OR = (2/1)/(2/1) = 1 = exp(-1.659e-15)
with(dd, table(disease[smoking==0], sex[smoking==0]))
```

```
##
##     0 1
## 0 1 1
## 1 2 2
```

The interaction effect is a ratio of two ORs. We
need the OR for sex among smokers.

```
## disease-sex distribution among smokers:
## OR = (2/2)/(0/2) = infinity
with(dd, table(disease[smoking==1], sex[smoking==1]))
```

```
##
##     0 1
## 0 2 2
## 1 0 2
```

The ratio of the last 2 ORs is infinity/1 =
infinity, which is estimated as exp(1.857e1) in
logistic regression.

Interaction Terms in Logistic Regression (cont'd)

Now consider an interaction model between adult age (a continuous variable with values way above 0) and smoking,

$$\log\left(\frac{p}{1-p}\right) = \beta_0 + \beta_1 \cdot \text{age} + \beta_2 \cdot \text{smoking} + \beta_3 \cdot \text{age} \cdot \text{smoking}$$

$$= \begin{cases} \beta_0 + \beta_1 \cdot \text{age}, & \text{non-smoker}; \\ (\beta_0 + \beta_2) + (\beta_1 + \beta_3) \cdot \text{age}, & \text{smoker}. \end{cases}$$

- e^{β_0} is the <u>odds</u> for non-smokers at 0 year old (**nonexistent!**).
- e^{β_1} is the <u>OR</u> for one year older vs. current age among non-smokers. (The OR for one year older vs. current age among smokers is $e^{\beta_1+\beta_3}$.)
- e^{β_2} is the <u>OR</u> for smoking among 0 year olds (**nonexistent!**) (The OR for smoking among x year olds is $e^{\beta_2+\beta_3 x}$.)
- e^{β_3} is a <u>ratio of ORs</u>:

$$e^{\beta_3} = \frac{\text{OR for one-year-older among smokers}}{\text{OR for one-year-older among non-smokers}} = \frac{\text{OR for smoking among } x+1 \text{ year olds}}{\text{OR for smoking among } x \text{ year olds}}.$$

Hypothesis Testing for Coefficents, Confidence Intervals

Testing for $H_0 : \beta_j = 0$ vs. $H_1 : \beta_j \neq 0$:

```
mod4 = glm(disease ~ age + smoking + sex, data=dd,
           family=binomial)
summary(mod4)$coef
```

```
##             Estimate Std. Error   z value  Pr(>|z|)
## (Intercept) -4.5030382  4.5615830 -0.9871657 0.3235614
## age          0.1449225  0.1386784  1.0450260 0.2960109
## smoking     -1.3479594  1.4248541 -0.9460333 0.3441316
## sex          1.3523022  1.4547410  0.9295828 0.3525872
```

The p-value is
- 0.296 for testing if the effect of age is zero
- 0.344 for testing if the effect of smoking is zero
- 0.353 for testing if the effect of sex is zero

Confidence intervals: `confint.default()` computes CIs assuming asymptotic normality.

```
## 95% CIs on the log-odds and log-OR scale
confint.default(mod4)
```

```
##                2.5 %     97.5 %
## (Intercept) -13.4435765 4.4375001
## age          -0.1268821 0.4167272
## smoking      -4.1406222 1.4447034
## sex          -1.4989379 4.2035422
```

```
## check mannually for age
c(0.1449225-1.96*0.1386784,0.1449225+1.96*0.1386784)
```

```
## [1] -0.1268872  0.4167322
```

```
## 95% CIs on the odds and OR scale
exp(confint.default(mod4))
```

```
##                 2.5 %     97.5 %
## (Intercept) 1.450537e-06 84.563278
## age          8.808375e-01  1.516989
## smoking      1.591295e-02  4.240594
## sex          2.233673e-01 66.922968
```

If the MASS package is installed, `confint()` gives CIs based on profile likelihood.

Hypothesis Testing when There is an Interaction Term

Let us revisit the interaction model we used before.

```
mod3 = glm(disease ~ smoking * sex, data=dd,
           family=binomial)
summary(mod3)$coef
```

```
##                 Estimate  Std. Error       z value   Pr(>|z|)
## (Intercept)  6.931472e-01  1.224745  5.659523e-01  0.5714262
## smoking     -1.925922e+01  4612.202148 -4.175709e-03  0.9966683
## sex         -1.659121e-15  1.732051 -9.578940e-16  1.0000000
## smoking:sex  1.856607e+01  4612.202419  4.025424e-03  0.9967882
```

To test for significance of the interaction term, the p-value is 0.9967882.

However, the two p-values above this number are not p-values for testing the "main effect".

- $p = 0.9966683$ is the p-value for testing if smoking is significant among women (the "baseline" sex when sex=0)
- $p = 1.000$ is the p-value for testing if sex is significant among non-smokers (the "baseline" for smoking, smoking=0)

ROC Curve and Its AUC

The **ROC curve** is a graphical display of the predictive performance for logistic regression models if the fitted probabilities are used to classify the subjects.

- For any $0 < t < 1$, we can define a classifier $\mathcal{C}_t$, in which a subject is classified as $\hat{Y} = 1$ if $\hat{p} \geq t$ and $\hat{Y} = 0$ if $\hat{p} < t$. Comparing $\hat{Y}$ with the observed Y, we can compute the sensitivity and specificity of $\mathcal{C}_t$. The ROC curve is a plot of sensitivity vs. $1-$ specificity (equivalently, power vs. type I error rate, or TPR vs. FPR) for all classifiers $\mathcal{C}_t$.

- If there are n observations, there are at most $n + 1$ classifiers and at most $n + 1$ points on the ROC curve. (Two classifiers are extreme ones: one predicting all subjects as 0 and the other predicting all subjects as 1, which correspond to $(0, 0)$ and $(1, 1)$ on the ROC curve).

The area-under-curve (**AUC**) of an ROC curve is a summary measure of the ROC curve. It is a measure of concondance between Y and $\hat{p}$. It is the same as the C-index (concordance index).

lrm() for Logistic Regression

The function `lrm()` in the rms R package can also be used to fit logistic regression. Make sure to run the `options(datadist= )` line so that we can run `summary()` on lrm objects,

```r
library(rms)
options(datadist=datadist(dd))
mod5 = lrm(disease ~ age + smoking + sex, data=dd)
summary(mod5)
```

```
##              Effects                  Response : disease
##
## Factor       Low  High Diff. Effect  S.E.    Lower 0.95 Upper 0.95
## age          26.5 35.5 9     1.30430 1.2481  -1.141900   3.7505
##  Odds Ratio  26.5 35.5 9     3.68510 NA       0.319200  42.5440
## smoking      0.0  1.0  1    -1.34800 1.4249  -4.140600   1.4447
##  Odds Ratio  0.0  1.0  1     0.25977 NA       0.015913   4.2406
## sex          0.0  1.0  1     1.35230 1.4547  -1.498900   4.2035
##  Odds Ratio  0.0  1.0  1     3.86630 NA       0.223370  66.9230
```

```r
mod5$coef
```

```
## Intercept       age    smoking       sex
## -4.5030378 0.1449225 -1.3479593 1.3523021
```

```r
mod5
```

```
## Logistic Regression Model
##
## lrm(formula = disease ~ age + smoking + sex, data = dd)
##
##                    Model Likelihood    Discrimination    Rank D
##                      Ratio Test           Indexes
## Obs        12       LR chi2    3.33   R2        0.323    C
## 0           6       d.f.          3   R2(3,12)  0.027    Dxy
## 1           6       Pr(> chi2) 0.3430 R2(3,9)   0.036    gamma
## max |deriv| 4e-06                     Brier     0.193    tau-a
##
##           Coef    S.E.   Wald Z Pr(>|Z|)
## Intercept -4.5030 4.5616 -0.99  0.3236
## age        0.1449 0.1387  1.05  0.2960
## smoking   -1.3480 1.4249 -0.95  0.3441
## sex        1.3523 1.4547  0.93  0.3526
##
```

```r
mod4 = glm(disease ~ age + smoking + sex, data=dd,
           family=binomial)
summary(mod4)$coef
```

```
##               Estimate Std. Error   z value  Pr(>|z|)
## (Intercept) -4.5030382  4.5615830 -0.9871657 0.3235614
## age          0.1449225  0.1386784  1.0450260 0.2960109
## smoking     -1.3479594  1.4248541 -0.9460333 0.3441316
## sex          1.3523022  1.4547410  0.9295828 0.3525872
```

polr() and orm() for Ordinal Logistic Regression

```
## Simulate a dataset with x and ordinal y
set.seed(20); n = 100
x = rnorm(n) + 10; y0 = 10 + (x-9)^2 + rnorm(n,0,2)
range(y0)
```

```
## [1]  6.443783 21.653317
```

```
y = cut(y0, breaks=c(0,9,12,15,100))
table(y)
```

```
## y
##    (0,9]   (9,12]  (12,15] (15,100]
##       16       34       32       18
```

In `polr()` from the MASS package, the model is $G_1(Pr(Y \leq y|X)) = \alpha_1 - \beta_1^T X$
(focus on lower side of every dichotomization)

```
library(MASS)
mod8 = polr(y ~ x)   ## default is logit link
summary(mod8)$coef
```

```
##                     Value Std. Error  t value
## x               1.665099   0.265234 6.277849
## (0,9]|(9,12]   14.268944   2.522839 5.655908
## (9,12]|(12,15] 16.553371   2.651118 6.243921
## (12,15]|(15,100] 18.887341 2.837238 6.656945
```

In `orm()` from the rms package, the model is
$G_2(Pr(Y \geq y|X)) = \alpha_2 + \beta_2^T X$
(focus on upper side of every dichotomization)

```
library(rms)
mod6 = orm(y ~ x)   ## default is logit link
mod6$coef
```

```
##   y>=(9,12] y>=(12,15] y>=(15,100]          x
## -14.268989 -16.553426 -18.887431   1.665106
```

```
## For logit link, lrm() can do it too
mod7 = lrm(y ~ x)
mod7$coef
```

```
##   y>=(9,12] y>=(12,15] y>=(15,100]          x
## -14.269011 -16.553449 -18.887460   1.665108
```

The models from `polr()` and `orm()` focus on different sides of the dichotomizations. They are equivalent when
$G_1(t) = -G_2(1 - t)$, $\alpha_1 = -\alpha_2$, $\beta_1 = \beta_2$

polr() and orm() for Ordinal Logistic Regression (cont'd)

G_1 (polr)	G_2 (orm)
logit	logit
probit	probit
cloglog	loglog
loglog	cloglog

```
polr.probit = polr(y ~ x, method="probit")
summary(polr.probit)$coef
```

```
##                    Value Std. Error  t value
## x             0.9273884  0.1358147 6.828337
## (0,9]|(9,12]   7.9180455  1.3035356 6.074284
## (9,12]|(12,15] 9.2071335  1.3506143 6.816997
## (12,15]|(15,100] 10.5779477 1.4426212 7.332450
```

```
orm.probit = orm(y ~ x, family="probit")
orm.probit$coef
```

```
##   y>=(9,12]  y>=(12,15] y>=(15,100]           x
## -7.9016602  -9.1894356 -10.5595105   0.9256388
```

```
polr.cloglog = polr(y ~ x, method="cloglog")
summary(polr.cloglog)$coef
```

```
##                    Value Std. Error  t value
## x             0.7972981  0.111467 7.152775
## (0,9]|(9,12]   5.9399438  1.053868 5.636325
## (9,12]|(12,15] 7.4955433  1.056307 7.095993
## (12,15]|(15,100] 8.8876164 1.169070 7.602298
```

```
orm.loglog = orm(y ~ x, family="loglog")
orm.loglog$coef
```

```
##   y>=(9,12]  y>=(12,15] y>=(15,100]           x
## -5.9399708  -7.4955694  -8.8876413   0.7973004
```

```
polr.loglog = polr(y ~ x, method="loglog")
summary(polr.loglog)$coef
```

```
##                    Value Std. Error  t value
## x             1.129081  0.1795183 6.289503
## (0,9]|(9,12]   10.316182  1.7631654 5.850944
## (9,12]|(12,15] 11.641869  1.8401061 6.326738
## (12,15]|(15,100] 13.267885 1.9360576 6.853043
```

```
orm.cloglog = orm(y ~ x, family="cloglog")
orm.cloglog$coef
```

```
##   y>=(9,12]  y>=(12,15] y>=(15,100]           x
## -10.31311   -11.63734   -13.26638    1.12872
```