# Log-linear Regression (practical)

Note updated April 13, 2019

Wan Nor Arifin
Unit of Biostatistics and Research Methodology,
Universiti Sains Malaysia.

*Email: wnarifin@usm.my*
*Website: wnarifin.github.io*



## Contents



## 1 Steps in log-linear modeling

Steps (Eye & Mun (2013), pg 81-84):

1. Specify models to be tested

2. Estimate the models: parameters, expected frequencies, residuals
3. Hypothesis testing: 1. Overall GOF 2. Parameters significance. 3. Standardized residuals (~z)
4. Model interpretation: model fit, parameters, ORs

# 2 Preliminaries

## 2.1 Load library

```
library(epiDisplay)  # to use poisgof
```

## 2.2 Data

Data in Eye & Mun (2013), pg. 162:

Table 1: Cross-classification table of victim, defendant and penalty.

| victim | defendant | penalty | freq |
|---|---|---|---|
| black | black | death_no | 593 |
| black | black | death_yes | 14 |
| black | white | death_no | 284 |
| black | white | death_yes | 38 |
| white | black | death_no | 25 |
| white | black | death_yes | 1 |
| white | white | death_no | 272 |
| white | white | death_yes | 23 |

```
penaltyTab = read.table(header = T, text = "
victim defendant  penalty freq
black     black death_no   593
black     black death_yes   14
black     white death_no   284
black     white death_yes   38
white     black death_no    25
white     black death_yes    1
white     white death_no   272
white     white death_yes   23")
penaltyTab
```

```
##   victim defendant   penalty freq
## 1  black     black  death_no  593
## 2  black     black death_yes   14
## 3  black     white  death_no  284
## 4  black     white death_yes   38
## 5  white     black  death_no   25
## 6  white     black death_yes    1
## 7  white     white  death_no  272
## 8  white     white death_yes   23
```

# 3 Step 1: Specify models to be tested

## 3.1 3 way table

3 variables, D, P, V

## 3.2 Fit hierarchical models

### 3.2.1 No interaction, Independence model:

1. D, P, V

### 3.2.2 One 2-way interaction

1. D, PV
2. P, DV
3. V, DP

### 3.2.3 Two 2-way interaction

1. DP, DV
2. DP, PV
3. DV, PV

### 3.2.4 Three 2-way interaction

1. DP, DV, PV

### 3.2.5 3-way interaction, Saturated model

1. DPV

In total 9 models.

# 4 Step 2 (Estimate the models) and 3 (Hypothesis testing)

These steps are together in results:

1. Step 2: Estimate the models.
   a. Parameters.
   b. Expected frequencies.
   c. Residuals.
2. Step 3: Hypothesis testing.
   a. Overall GOF.
   b. Parameters significance.
   c. Standardized residuals (~z).

## 4.1 GOF, Model-model comparison

List down X2, G2 for GOF, AIC in table form.

Model comparisons:

1. List down G2, model_n vs model_n-1; G2_n-1 - G2_n -> LR Test
2. List down delta_AIC, AIC_n - AIC_n-1

## 4.2 No interaction, Independence model

```r
# D, P, V
ll.model0 = glm(freq ~ defendant + penalty + victim, data = penaltyTab, family = poisson)  # Step 2
summary(ll.model0)  # Step 2 & 3
```

```
##
## Call:
## glm(formula = freq ~ defendant + penalty + victim, family = poisson,
##     data = penaltyTab)
##
## Deviance Residuals:
##       1        2        3        4        5        6        7        8
##   6.830   -3.033   -7.540    1.815  -12.840   -3.631    9.040    3.647
##
## Coefficients:
##                  Estimate Std. Error z value Pr(>|z|)
## (Intercept)       6.09095    0.04368 139.444   <2e-16 ***
## defendantwhite   -0.02560    0.05657  -0.453    0.651
## penaltydeath_yes -2.73744    0.11836 -23.128   <2e-16 ***
## victimwhite      -1.06267    0.06474 -16.414   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for poisson family taken to be 1)
##
##     Null deviance: 1857.90  on 7  degrees of freedom
## Residual deviance:  389.07  on 4  degrees of freedom
## AIC: 442.24
##
## Number of Fisher Scoring iterations: 5
```

```r
names(ll.model0)  # detailed results options
```

```
##  [1] "coefficients"      "residuals"         "fitted.values"     "effects"
##  [5] "R"                 "rank"              "qr"                "family"
##  [9] "linear.predictors" "deviance"          "aic"               "null.deviance"
## [13] "iter"              "weights"           "prior.weights"     "df.residual"
## [17] "df.null"           "y"                 "converged"         "boundary"
## [21] "model"             "call"              "formula"           "terms"
## [25] "data"              "offset"            "control"           "method"
## [29] "contrasts"         "xlevels"
```

```r
# Residuals, obs vs pred
penaltyTab$pred = ll.model0$fitted.values
penaltyTab
```

```
##   victim defendant   penalty freq       pred
## 1  black     black  death_no  593 441.842508
## 2  black     black death_yes   14  28.603092
## 3  black     white  death_no  284 430.674292
## 4  black     white death_yes   38  27.880108
## 5  white     black  death_no   25 152.671093
## 6  white     black death_yes    1   9.883308
## 7  white     white  death_no  272 148.812108
## 8  white     white death_yes   23   9.633492
```

```r
cbind(penaltyTab, rawres = with(penaltyTab, freq - pred), stdres = rstandard(ll.model0, type = "pearson
```

```
##   victim defendant   penalty freq       pred      rawres     stdres
## 1  black     black  death_no  593 441.842508  151.157492  18.150369
## 2  black     black death_yes   14  28.603092  -14.603092  -3.544471
## 3  black     white  death_no  284 430.674292 -146.674292 -17.634181
## 4  black     white death_yes   38  27.880108   10.119892   2.469017
## 5  white     black  death_no   25 152.671093 -127.671093 -16.390236
## 6  white     black death_yes    1   9.883308   -8.883308  -3.084245
## 7  white     white  death_no  272 148.812108  123.187892  15.837421
## 8  white     white death_yes   23   9.633492   13.366508   4.690347
```

```r
## stdres = standardized pearson residual: Agresti 3.13, pg81 GOF
poisgof(ll.model0)  # G2
```

```
## $results
## [1] "Goodness-of-fit test for Poisson assumption"
##
## $chisq
## [1] 389.0726
##
## $df
## [1] 4
##
## $p.value
## [1] 6.385409e-83
```

```r
penaltyTab$x2 = with(penaltyTab, (freq - pred)^2/pred)
# X2 formula: von Eye & Mun, pg25; Agresti 3.10, pg79
list(results = "X2 GOF", chisq = sum(penaltyTab$x2), df = ll.model0$df.residual, p.value = pchisq(sum(pe
    ll.model0$df.residual, lower.tail = F))  # X2
```

```
## $results
## [1] "X2 GOF"
##
## $chisq
## [1] 348.065
##
## $df
## [1] 4
##
## $p.value
## [1] 4.589572e-74
```

```r
AIC(ll.model0)
```

```
## [1] 442.2404
```

```r
# Parameters
cbind(round(summary(ll.model0)$coefficients, 3), round(confint(ll.model0), 3))
```

```
## Waiting for profiling to be done...

##                   Estimate Std. Error z value Pr(>|z|)  2.5 % 97.5 %
## (Intercept)          6.091      0.044 139.444    0.000  6.004  6.176
## defendantwhite      -0.026      0.057  -0.453    0.651 -0.137  0.085
## penaltydeath_yes    -2.737      0.118 -23.128    0.000 -2.978 -2.513
## victimwhite         -1.063      0.065 -16.414    0.000 -1.191 -0.937
```

```r
idr.display(ll.model0)  # ORs
```

```
##
## Poisson regression predicting freq
##
##                                   crude IDR(95%CI)  adj. IDR(95%CI)   P(Wald's test)
## defendant: white vs black         0.97 (0.87,1.09)  0.97 (0.87,1.09)  0.651
##
## penalty: death_yes vs death_no 0.06 (0.05,0.08)  0.06 (0.05,0.08)  < 0.001
##
## victim: white vs black            0.35 (0.3,0.39)   0.35 (0.3,0.39)   < 0.001
##
##                                   P(LR-test)
## defendant: white vs black         0.651
##
## penalty: death_yes vs death_no < 0.001
##
## victim: white vs black            < 0.001
##
## Log-likelihood = -217.1202
## No. of observations = 8
## AIC value = 442.2404
```

```r
# Model-model comparison
AIC(ll.model0) - AIC(ll.model_ <- glm(freq ~ 1, data = penaltyTab, family = poisson))  # vs empty model
```

```
## [1] -1462.825
```

```r
anova(ll.model_, ll.model0, test = "LRT")  # vs empty model
```

```
## Analysis of Deviance Table
##
## Model 1: freq ~ 1
## Model 2: freq ~ defendant + penalty + victim
##   Resid. Df Resid. Dev Df Deviance  Pr(>Chi)
## 1         7    1857.90
## 2         4     389.07  3   1468.8 < 2.2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

### 4.3 One 2-way interaction

```r
# D, PV
ll.model1 = glm(freq ~ defendant + penalty * victim, data = penaltyTab, family = poisson)  # Step 2
```

```r
summary(ll.model1)
```

```
##
## Call:
## glm(formula = freq ~ defendant + penalty * victim, family = poisson,
##     data = penaltyTab)
##
## Deviance Residuals:
##       1        2        3        4        5        6        7        8
##   6.717   -2.641   -7.639    2.270  -12.692   -4.161    9.244    2.866
##
## Coefficients:
##                             Estimate Std. Error z value Pr(>|z|)
## (Intercept)                  6.09608    0.04382 139.123   <2e-16 ***
## defendantwhite              -0.02560    0.05657  -0.453    0.651
## penaltydeath_yes            -2.82526    0.14273 -19.795   <2e-16 ***
## victimwhite                 -1.08277    0.06714 -16.128   <2e-16 ***
## penaltydeath_yes:victimwhite  0.30958    0.25573   1.211    0.226
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for poisson family taken to be 1)
##
##     Null deviance: 1857.90  on 7  degrees of freedom
## Residual deviance:  387.66  on 3  degrees of freedom
## AIC: 442.83
##
## Number of Fisher Scoring iterations: 5
```

```r
poisgof(ll.model1)
```

```
## $results
## [1] "Goodness-of-fit test for Poisson assumption"
##
## $chisq
## [1] 387.658
##
## $df
## [1] 3
##
## $p.value
## [1] 1.043321e-83
```

```r
penaltyTab$pred = ll.model1$fitted.values
penaltyTab
```

```
##   victim defendant   penalty freq     pred          x2
## 1  black     black  death_no  593 444.1128  51.712063
## 2  black     black death_yes   14  26.3328   7.455498
## 3  black     white  death_no  284 432.8872  49.952710
## 4  black     white death_yes   38  25.6672   3.673308
## 5  white     black  death_no   25 150.4008 106.764860
## 6  white     black death_yes    1  12.1536   7.984488
## 7  white     white  death_no  272 146.5992 101.975955
## 8  white     white death_yes   23  11.8464  18.546080
```

```r
penaltyTab$x2 = with(penaltyTab, (freq - pred)^2/pred)
list(results = "X2 GOF", chisq = sum(penaltyTab$x2), df = ll.model1$df.residual, p.value = pchisq(sum(pe
    ll.model1$df.residual, lower.tail = F))
```

```
## $results
## [1] "X2 GOF"
##
## $chisq
## [1] 345.3852
##
## $df
## [1] 3
##
## $p.value
## [1] 1.489037e-74
```

```r
AIC(ll.model1)
```

```
## [1] 442.8258
```

```r
anova(ll.model0, ll.model1, test = "LRT")
```

```
## Analysis of Deviance Table
##
## Model 1: freq ~ defendant + penalty + victim
## Model 2: freq ~ defendant + penalty * victim
##   Resid. Df Resid. Dev Df Deviance Pr(>Chi)
## 1         4     389.07
## 2         3     387.66  1   1.4146   0.2343
```

Now repeat same commands for all models…

```r
# P, DV V, DP
```

### 4.4 Two 2-way interaction

```r
# DP, DV DP, PV DV, PV
```

### 4.5 Three 2-way interaction

```r
# DP, DV, PV
```

### 4.6 3-way interaction, Saturated model

```r
# DPV
```

### 4.7 Write model fit results in a table

Table 2: Model-model comparison of GOF.

| Model | $G^2$ | $\chi^2$ | df | $P$-value ($G^2$) | AIC |
|---|---|---|---|---|---|
| (D, P, V) | 389.07 | 348.07 | 4 | $< 0.001$ | 442.24 |
| (D, PV) | 387.66 | 345.39 | 3 | $< 0.001$ | 442.83 |
| (models) | - | - | - | - | - |

# 5 Step 4: Model interpretation; model fit, parameters, ORs

## 5.1 Interpret your results

Which model has best model fit? Interpret the chosen model.

# References


Agresti, A. (2002). *Categorical data analysis (2nd ed.* Hoboken, New Jersey: John Wiley & Sons.

Chongsuvivatwong, V. (2018). *EpiDisplay: Epidemiological data display package.* Retrieved from https://CRAN.R-project.org/package=epiDisplay

Eye, A. von, & Mun, E. (2013). *Log-linear modeling: Concepts, interpretation, and application.* Hoboken, New Jersey: John Wiley & Sons.

Stevens, J. P. (2009). *Applied multivariate statistics for the social sciences (5th eds.).* New York: Taylor & Francis Group.