# #03_Poisson対応

harada

2024-09-01

```
summary(cars)
```

```
##      speed           dist
##  Min.   : 4.0   Min.   :  2.00
##  1st Qu.:12.0   1st Qu.: 26.00
##  Median :15.0   Median : 36.00
##  Mean   :15.4   Mean   : 42.98
##  3rd Qu.:19.0   3rd Qu.: 56.00
##  Max.   :25.0   Max.   :120.00
```

##dataよみこみ

```
Dataset<- read.csv('https://raw.githubusercontent.com/harabou/Biostat_Kyoto_pref/main/data/%2303/earinfection.csv')
```

##Poisson

```
GLM.1 <- glm(infections ~ age + gender + location + swimmer, family=poisson(log), data=Dataset)
summary(GLM.1)
```

```
##
## Call:
## glm(formula = infections ~ age + gender + location + swimmer,
##     family = poisson(log), data = Dataset)
##
## Coefficients:
##                 Estimate Std. Error z value Pr(>|z|)
## (Intercept)     -0.12261    0.13706  -0.895  0.37100
## age20-24        -0.37442    0.12836  -2.917  0.00354 **
## age25-29        -0.18973    0.13009  -1.458  0.14473
## genderMale      -0.08985    0.11231  -0.800  0.42371
## locationNonBeach  0.53454    0.10668   5.011 5.43e-07 ***
## swimmerOccas      0.61149    0.10500   5.823 5.77e-09 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for poisson family taken to be 1)
##
##     Null deviance: 824.51  on 286  degrees of freedom
## Residual deviance: 755.43  on 281  degrees of freedom
## AIC: 1139.8
##
## Number of Fisher Scoring iterations: 6
```

```
exp(coef(GLM.1))  # Exponentiated coefficients
```

```
##      (Intercept)         age20-24         age25-29       genderMale
##        0.8846066        0.6876857        0.8271832        0.9140681
## locationNonBeach     swimmerOccas
##        1.7066565        1.8431689
```

# Overdispersion

```
dp <- sum(residuals (GLM.1,type="pearson")^2) / GLM.1$df.res
summary(GLM.1, dispersion=dp)
```

```
##
## Call:
## glm(formula = infections ~ age + gender + location + swimmer,
##     family = poisson(log), data = Dataset)
##
## Coefficients:
##                  Estimate Std. Error z value Pr(>|z|)
## (Intercept)      -0.12261    0.25197  -0.487  0.62653
## age20-24         -0.37442    0.23599  -1.587  0.11260
## age25-29         -0.18973    0.23917  -0.793  0.42762
## genderMale       -0.08985    0.20648  -0.435  0.66345
## locationNonBeach  0.53454    0.19613   2.725  0.00642 **
## swimmerOccas      0.61149    0.19304   3.168  0.00154 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for poisson family taken to be 3.379853)
##
##     Null deviance: 824.51  on 286  degrees of freedom
## Residual deviance: 755.43  on 281  degrees of freedom
## AIC: 1139.8
##
## Number of Fisher Scoring iterations: 6
```

##negative binominal

```
library(MASS)
result_nb<-glm.nb(infections~age + gender + location + swimmer, data=Dataset)
summary(result_nb)
```

```
##
## Call:
## glm.nb(formula = infections ~ age + gender + location + swimmer,
##     data = Dataset, init.theta = 0.5759778332, link = log)
##
## Coefficients:
##                 Estimate Std. Error z value Pr(>|z|)
## (Intercept)     -0.06082    0.23914  -0.254  0.79924
## age20-24        -0.42924    0.23433  -1.832  0.06698 .
## age25-29        -0.26250    0.24050  -1.091  0.27506
## genderMale      -0.14002    0.20789  -0.674  0.50063
## locationNonBeach  0.54554    0.19605   2.783  0.00539 **
## swimmerOccas      0.60305    0.18965   3.180  0.00147 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for Negative Binomial(0.576) family taken to be 1)
##
##     Null deviance: 289.90  on 286  degrees of freedom
## Residual deviance: 269.13  on 281  degrees of freedom
## AIC: 904.69
##
## Number of Fisher Scoring iterations: 1
##
##
##               Theta:  0.5760
##           Std. Err.:  0.0903
##
##  2 x log-likelihood:  -890.6900
```

```
exp(coef(result_nb))  # Exponentiated coefficients
```

```
##      (Intercept)         age20-24         age25-29       genderMale
##        0.9409927        0.6510050        0.7691229        0.8693440
## locationNonBeach     swimmerOccas
##        1.7255466        1.8276788
```

##Zero-inflated Poisson

```
library(pscl)
```

```
## Warning: パッケージ 'pscl' はバージョン 4.3.3 の R の下で造られました
```

```
## Classes and Methods for R originally developed in the
## Political Science Computational Laboratory
## Department of Political Science
## Stanford University (2002-2015),
## by and under the direction of Simon Jackman.
## hurdle and zeroinfl functions by Achim Zeileis.
```

```
result_zi <-  zeroinfl(infections~age + gender + location + swimmer, data=Dataset)
summary(result_zi)
```

```
##
## Call:
## zeroinfl(formula = infections ~ age + gender + location + swimmer, data = Dataset)
##
## Pearson residuals:
##     Min      1Q  Median      3Q     Max
## -1.0340 -0.7384 -0.5824  0.4715  8.3272
##
## Count model coefficients (poisson with log link):
##                  Estimate Std. Error z value Pr(>|z|)
## (Intercept)        0.8113     0.1507   5.384 7.29e-08 ***
## age20-24          -0.3825     0.1507  -2.538   0.0111 *
## age25-29          -0.2505     0.1462  -1.714   0.0866 .
## genderMale        -0.1464     0.1348  -1.086   0.2774
## locationNonBeach   0.1887     0.1270   1.486   0.1373
## swimmerOccas       0.5147     0.1235   4.169 3.06e-05 ***
##
## Zero-inflation model coefficients (binomial with logit link):
##                   Estimate Std. Error z value Pr(>|z|)
## (Intercept)       0.381059   0.334142   1.140   0.2541
## age20-24         -0.004427   0.344180  -0.013   0.9897
## age25-29         -0.024362   0.345686  -0.070   0.9438
## genderMale       -0.036735   0.300468  -0.122   0.9027
## locationNonBeach -0.725336   0.284673  -2.548   0.0108 *
## swimmerOccas     -0.189300   0.277742  -0.682   0.4955
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Number of iterations in BFGS optimization: 18
## Log-likelihood: -471.1 on 12 Df
```

##Zero-inflated negative binominal

```
result_zinb <- zeroinfl(infections~age + gender + location + swimmer, data=Dataset,dist="negbin")
summary(result_zinb)
```

```
##
## Call:
## zeroinfl(formula = infections ~ age + gender + location + swimmer, data = Dataset,
##     dist = "negbin")
##
## Pearson residuals:
##     Min      1Q  Median      3Q     Max
## -0.7809 -0.6347 -0.4571  0.3936  5.7963
##
## Count model coefficients (negbin with log link):
##                  Estimate Std. Error z value Pr(>|z|)
## (Intercept)       0.09276    0.27261   0.340 0.733654
## age20-24         -0.32053    0.23499  -1.364 0.172562
## age25-29         -0.17129    0.26573  -0.645 0.519182
## genderMale        0.07730    0.22721   0.340 0.733705
## locationNonBeach  0.12344    0.24796   0.498 0.618615
## swimmerOccas      0.69295    0.19894   3.483 0.000495 ***
## Log(theta)       -0.23651    0.21516  -1.099 0.271665
##
## Zero-inflation model coefficients (binomial with logit link):
##                  Estimate Std. Error z value Pr(>|z|)
## (Intercept)       -2.3466     1.7221  -1.363    0.173
## age20-24           0.5716     1.0841   0.527    0.598
## age25-29           0.4764     0.9716   0.490    0.624
## genderMale         1.2528     1.1912   1.052    0.293
## locationNonBeach -12.9315   323.6701  -0.040    0.968
## swimmerOccas       0.7042     0.8837   0.797    0.426
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Theta = 0.7894
## Number of iterations in BFGS optimization: 56
## Log-likelihood: -440.9 on 13 Df
```

#AIC比較

```
c(AIC.model1=AIC(GLM.1), AIC.model2=AIC(result_nb),AIC.model3=AIC(result_zi), AIC.model4=AIC(result_zi
nb))
```

```
## AIC.model1 AIC.model2 AIC.model3 AIC.model4
##  1139.8280   904.6900   966.2640   907.8795
```