Glm2

2023-10-28

Bihan Qian

```
#import data
library(readxl)
```

```
## Warning: package 'readxl' was built under R version 4.2.3
```

```
valve <- read_excel("C:/Users/11139/Desktop/STAT5391/valve.xlsx")
```

1(a) Code

1. Consider the Valve characteristics data (Display on next Slide, Ramsey and Schafer).

a. Using an appropriate Poisson model, determine if there is association between valve failure and operator

```
print("When failure ~ system + operator + valve + size + mode")
```

```
## [1] "When failure ~ system + operator + valve + size + mode"
```

```
mode1 <- glm(failure~system+operator+valve+size+mode, data=valve, offset=log(time), family="pois
son")
summary(mode1)
```

```
##
## Call:
## glm(formula = failure ~ system + operator + valve + size + mode,
##     family = "poisson", data = valve, offset = log(time))
##
## Deviance Residuals:
##     Min       1Q   Median       3Q      Max
## -1.9963  -1.0531  -0.5743   0.9912   2.2809
##
## Coefficients:
##             Estimate Std. Error z value Pr(>|z|)
## (Intercept)  -6.4725     1.9826  -3.265 0.001096 **
## system        0.5685     0.2045   2.780 0.005429 **
## operator     -0.9857     0.2332  -4.228 2.36e-05 ***
## valve         0.9627     0.3348   2.875 0.004040 **
## size          2.0272     0.3383   5.992 2.08e-09 ***
## mode         -1.0435     0.3166  -3.296 0.000982 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for poisson family taken to be 1)
##
##     Null deviance: 98.883  on 25  degrees of freedom
## Residual deviance: 46.363  on 20  degrees of freedom
## AIC: 99.649
##
## Number of Fisher Scoring iterations: 5
```

```r
# we assume operator: all possible columns

print("When failure ~ operator")
```

```
## [1] "When failure ~ operator"
```

```r
mode2 <- glm(failure~operator, data=valve, offset=log(time), family="poisson")
# we assume in the question "operator" is: column "operator"
summary(mode2)
```

```
##
## Call:
## glm(formula = failure ~ operator, family = "poisson", data = valve,
##     offset = log(time))
##
## Deviance Residuals:
##     Min       1Q   Median       3Q      Max
## -2.5190  -1.6817  -0.7372   0.2713   6.7122
##
## Coefficients:
##             Estimate Std. Error z value Pr(>|z|)
## (Intercept)  -0.1050     0.3997  -0.263    0.793
## operator     -0.2288     0.1434  -1.596    0.111
##
## (Dispersion parameter for poisson family taken to be 1)
##
##     Null deviance: 98.883  on 25  degrees of freedom
## Residual deviance: 96.465  on 24  degrees of freedom
## AIC: 141.75
##
## Number of Fisher Scoring iterations: 6
```

1(a) Answer

When failure ~ system + operator + valve + size + mode:

- As p-value 2.36e-05 for operator, we can see a statistical significant association between failure and operator.

When failure ~ operator:

- As p-value 0.111 for operator, we cannot see a statistical significant association between failure and operator.

2(a) Code

2. In the above:

a. Interpret the estimated parameters

```
print("When failure ~ system + operator + valve + size + mode")
```

```
## [1] "When failure ~ system + operator + valve + size + mode"
```

```
exp(-0.9857)
```

```
## [1] 0.3731779
```

```
exp(-6.4725-0.9857*1)
```

```
## [1] 0.0005766933
```

```
exp(-6.4725-0.9857*2) #operator2
```

```
## [1] 0.0002152092
```

```
exp(-6.4725-0.9857*3) #operator3
```

```
## [1] 8.031132e-05
```

```
exp(-6.4725-0.9857*4) #operator4
```

```
## [1] 2.997041e-05
```

2(a) Answer

When failure ~ system + operator + valve + size + mode:

- As p-value 2.36e-05 for operator, we can see a statistical significant association between failure and operator.
- Y is number of failure.
- When operator=1, the mean number of failure is exp(-6.4725-0.9857*1) as 0.0005766933.
- When operator=2, the mean number of failure is exp(-6.4725-0.9857*2) as 0.0002152092.
- When operator=3, the mean number of failure is exp(-6.4725-0.9857*3) as 8.031132e-05.
- When operator=4, the mean number of failure is exp(-6.4725-0.9857*4) as 2.997041e-05.

```
print("When failure ~ operator")
```

```
## [1] "When failure ~ operator"
```

```
exp(-0.9857)
```

```
## [1] 0.3731779
```

```
exp(-0.1050-0.2288*1)
```

```
## [1] 0.716197
```

```
exp(-0.1050-0.2288*2) #operator2
```

```
## [1] 0.5697258
```

```
exp(-0.1050-0.2288*3) #operator3
```

```
## [1] 0.4532099
```

```
exp(-0.1050-0.2288*4) #operator4
```

```
## [1] 0.3605228
```

When failure ~ operator: In another case, if we only consider “operator” (failure~operator)in poisson:

- P-value 0.111 suggesting that there is no strong evidence to conclude that “operator” has a significant impact on the number of failures in this Poisson regression model.
- When operator=1, the mean number of failure is exp(-0.1050-0.2288*1) as 0.716197. Operator 1* reduce the mean number of failure by (1-0.716197)x100%.
- When operator=2, the mean number of failure is exp(-0.1050-0.2288*2) as 0.5697258.Operator 1 reduce the mean number of failure by (1-0.5697258)x100%.
- When operator=3, the mean number of failure is exp(-0.1050-0.2288*3) as 0.4532099.Operator 1 reduce the mean number of failure by (1-0.4532099)x100%.
- When operator=4, the mean number of failure is exp(-0.1050-0.2288*4) as 0.3605228.Operator 1 reduce the mean number of failure by (1-0.3605228)x100%.

2(b) Code

b. Assess the goodness of fit of the model

```
print("When failure ~ system + operator + valve + size + mode")
```

```
## [1] "When failure ~ system + operator + valve + size + mode"
```

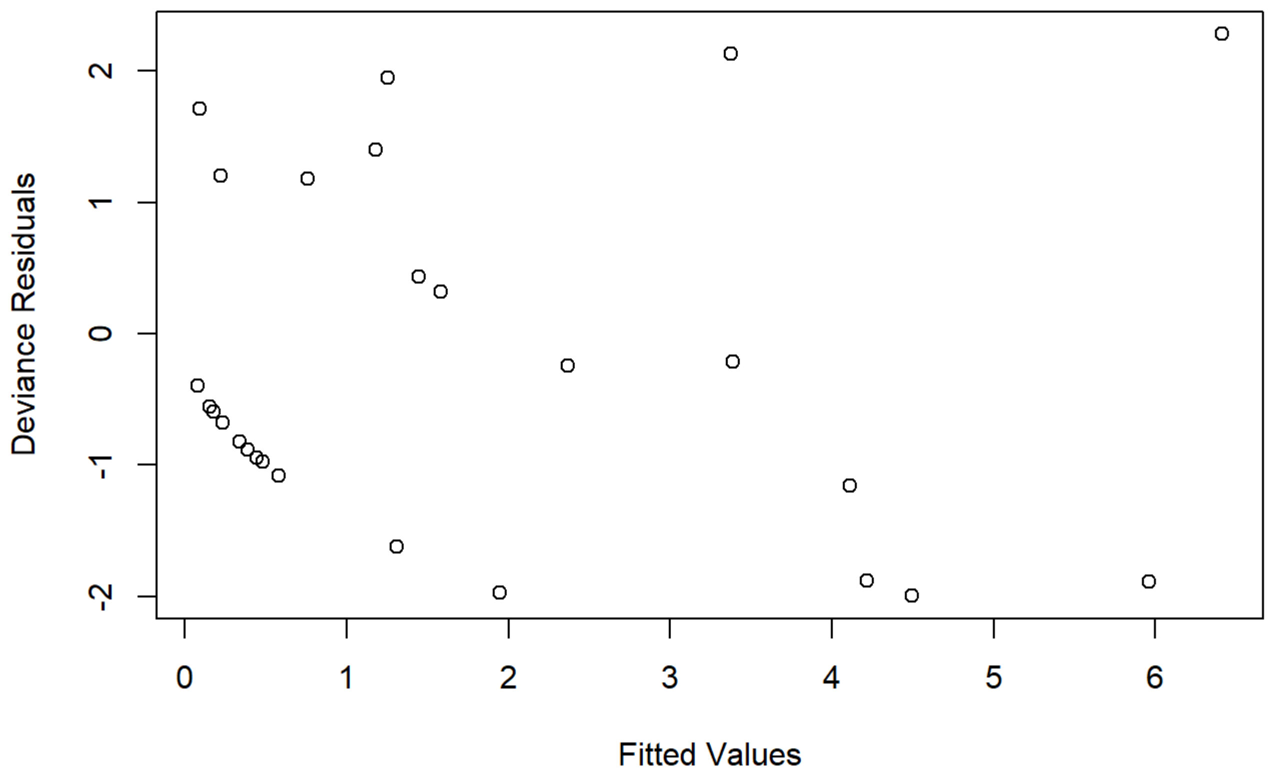
```
anova(mode1, test = "Chisq")
```

```
## Analysis of Deviance Table
##
## Model: poisson, link: log
##
## Response: failure
##
## Terms added sequentially (first to last)
##
##
##          Df Deviance Resid. Df Resid. Dev  Pr(>Chi)
## NULL                        25     98.883
## system    1    0.011        24     98.873 0.9182981
## operator  1    2.816        23     96.056 0.0933079 .
## valve     1    0.045        22     96.011 0.8320431
## size      1   38.209        21     57.803 6.357e-10 ***
## mode      1   11.439        20     46.363 0.0007191 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
with(mode1, cbind(res.deviance = deviance, df = df.residual,
  p = pchisq(deviance, df.residual, lower.tail=FALSE)))
```

```
##      res.deviance df            p
## [1,]     46.36335 20 0.0007184257
```

```
# resiudal plot
residuals <- residuals(mode1, type = "deviance")
plot(fitted(mode1), residuals, xlab = "Fitted Values", ylab = "Deviance Residuals")
```




```
print("When failure ~ operator")
```

```
## [1] "When failure ~ operator"
```

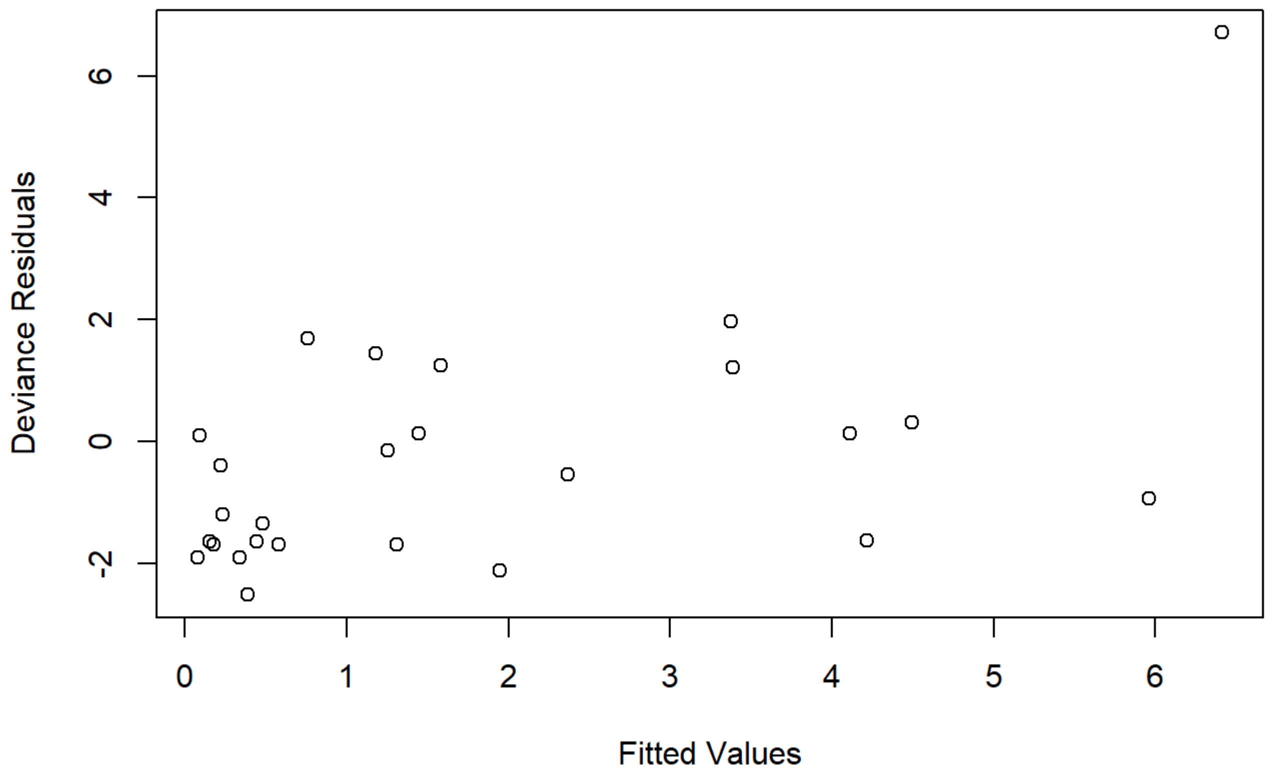
```
anova(mode2, test = "Chisq")
```

```
## Analysis of Deviance Table
##
## Model: poisson, link: log
##
## Response: failure
##
## Terms added sequentially (first to last)
##
##
##          Df Deviance Resid. Df Resid. Dev Pr(>Chi)
## NULL                        25     98.883
## operator  1   2.4181        24     96.465   0.1199
```

```
with(mode2, cbind(res.deviance = deviance, df = df.residual,
  p = pchisq(deviance, df.residual, lower.tail=FALSE)))
```

```
##      res.deviance df            p
## [1,]     96.46491 24 1.195876e-10
```

```
residuals <- residuals(mode2, type = "deviance")
plot(fitted(mode1), residuals, xlab = "Fitted Values", ylab = "Deviance Residuals")
```



#

2(b) Answer

When failure ~ system + operator + valve + size + mode:

- Operator p-value as 0.0933079, suggesting a marginal improvement in the model's goodness of fit, but it is not statistically significant. "size" and mode" are significant predictors, and their inclusion in the model greatly improves the model's goodness of fit.
- From residual plot and analysis of deviance as p-value 0.0007184257, the model does not fit data well with significant difference between predicted and observed.

When failure ~ operator:

- In another case, * Operator p-value as 0.0933079, operator is not a significant predictor. From residual plot and analysis of deviance as p-value 1.195876e-10, the model does not fit data well with significant difference between predicted and observed.

#3 Code 3. Repeat 1(a) using the glmnet package and comment on the results.

```
print("When failure ~ system + operator + valve + size + mode")
```

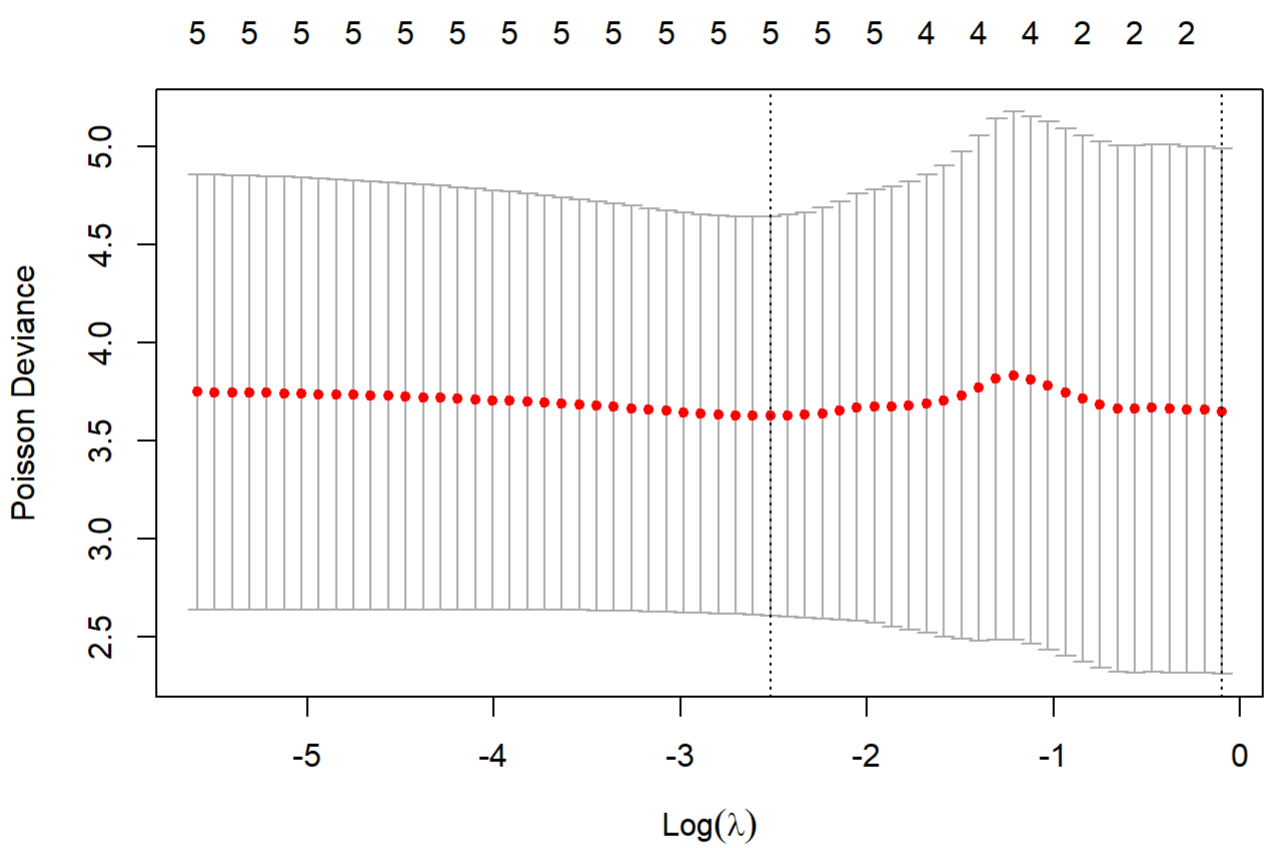
```
## [1] "When failure ~ system + operator + valve + size + mode"
```

```r
library(Matrix)
library(glmnet)
```

```
## Warning: package 'glmnet' was built under R version 4.2.3
```

```
## Loaded glmnet 4.1-8
```

```r
x <- model.matrix(failure ~ system + operator + valve + size + mode, data = valve)
y <- valve$failure
model_glmnet <- glmnet(x, y, family = "poisson")
cv_model <- cv.glmnet(x, y, family = "poisson",grouped=FALSE)
plot(cv_model)
```




```r
coef(cv_model, s = "lambda.min")
```

```
## 7 x 1 sparse Matrix of class "dgCMatrix"
##                      s1
## (Intercept) -3.0723214
## (Intercept)  .
## system       0.3493691
## operator    -0.1382686
## valve        0.5832666
## size         1.0390576
## mode        -1.0201250
```

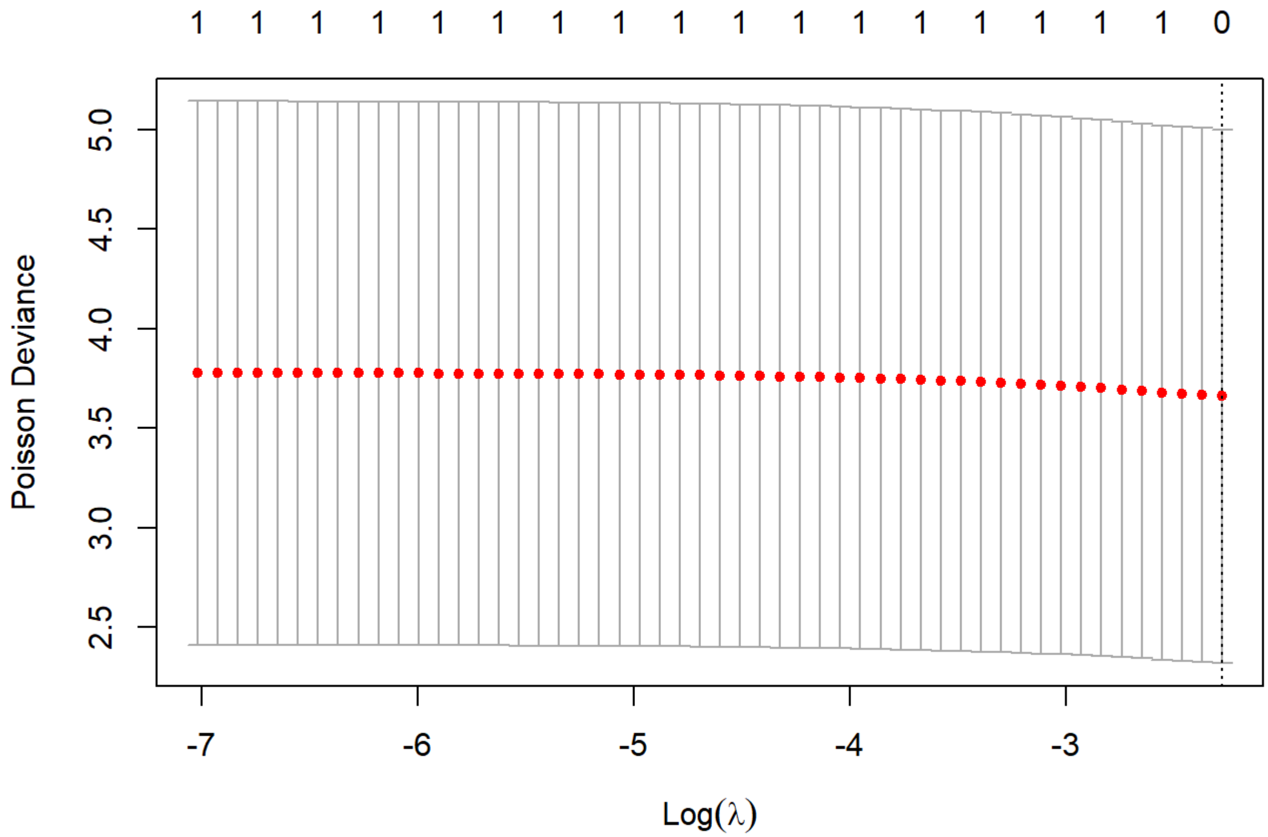
```
print("When failure ~ operator")
```

```
## [1] "When failure ~ operator"
```

```
x <- model.matrix(failure ~ operator, data = valve)
y <- valve$failure
cv_model <- cv.glmnet(x, y, family = "poisson")
```

```
## Warning: Option grouped=FALSE enforced in cv.glmnet, since < 3 observations per
## fold
```

```
plot(cv_model)
```




```
coef(cv_model, s = "lambda.min")
```

```
## 3 x 1 sparse Matrix of class "dgCMatrix"
##                    s1
## (Intercept) 0.5920511
## (Intercept) .
## operator    .
```

3 Answer

When failure ~ system + operator + valve + size + mode:

- The first intercept, represents the baseline rate of the event when all other predictor variables (system, operator, valve, size, mode) are set to zero, while accounting for the offset variable log(time). The second intercept count without the offset term log(time). We will use the first intercept.
- “s1” indicates failure=1. Size and mode are associated with failure while adjusting for offset term log(time).

When failure ~ operator:

- In another case, if we only consider “operator” (failure~operator)in poisson, operator is not significant associated with failure.