# 5291 hw8

Yijin Wang

11/16/2020

1.a

```
library(dplyr)
```

```
##
## Attaching package: 'dplyr'
```

```
## The following objects are masked from 'package:stats':
##
##     filter, lag
```

```
## The following objects are masked from 'package:base':
##
##     intersect, setdiff, setequal, union
```

```
#install.packages("Sleuth3")
data<-Sleuth3::ex2224
data<-data%>%mutate(System=as.factor(System),
                    Operator=as.factor(Operator),
                    Valve=as.factor(Valve),
                    Size=as.factor(Size),
                    Mode=as.factor(Mode))
head(data)
```

```
##   System Operator Valve Size Mode Failures Time
## 1      1        3     4    3    1        2    4
## 2      1        3     4    3    2        2    4
## 3      1        3     5    1    1        1    2
## 4      2        1     2    2    2        0    2
## 5      2        1     3    2    1        0    2
## 6      2        1     3    2    2        0    1
```

```
poi<-glm(Failures~Operator, family="poisson", data = data)
summary(poi)
```

```
##
## Call:
## glm(formula = Failures ~ Operator, family = "poisson", data = data)
##
## Deviance Residuals:
##     Min       1Q   Median       3Q      Max
## -2.0953  -1.9954  -0.9043   0.4427   8.1521
##
## Coefficients:
##             Estimate Std. Error z value Pr(>|z|)
## (Intercept)   0.7862     0.1054   7.459 8.72e-14 ***
```

```
## Operator2    -0.7862     0.5110  -1.539  0.12389
## Operator3    -0.4233     0.1812  -2.336  0.01951 *
## Operator4    -1.7417     0.4595  -3.791  0.00015 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for poisson family taken to be 1)
##
##     Null deviance: 413.76  on 89  degrees of freedom
## Residual deviance: 386.71  on 86  degrees of freedom
## AIC: 499.05
##
## Number of Fisher Scoring iterations: 6
```

P-values for operator1, 3, 4 are below 0.05, which reject the null hypothesis. Thus we conclude that there is association between valve failure and operator.

1.b

```
poi2<-glm(Failures~System+Operator+Valve+Size+Mode, family="poisson", offset = log(Time), data=data)
summary(poi2)
```

```
##
## Call:
## glm(formula = Failures ~ System + Operator + Valve + Size + Mode,
##     family = "poisson", data = data, offset = log(Time))
##
## Deviance Residuals:
##     Min       1Q   Median       3Q      Max
## -3.1892  -1.0074  -0.4357   0.3361   5.3138
##
## Coefficients:
##             Estimate Std. Error z value Pr(>|z|)
## (Intercept) -3.76867    0.81935  -4.600 4.23e-06 ***
## System2      0.91556    0.53184   1.721  0.08516 .
## System3      1.01881    0.50548   2.016  0.04385 *
## System4      1.22309    0.55518   2.203  0.02759 *
## System5      0.33292    0.58408   0.570  0.56869
## Operator2    0.70437    0.56669   1.243  0.21389
## Operator3   -1.19261    0.24851  -4.799 1.59e-06 ***
## Operator4   -2.47233    0.47660  -5.187 2.13e-07 ***
## Valve2       0.18533    0.76105   0.244  0.80761
## Valve3       0.60674    0.78107   0.777  0.43727
## Valve4       2.95894    0.60010   4.931 8.19e-07 ***
## Valve5       1.79318    0.61040   2.938  0.00331 **
## Valve6       1.00891    0.93009   1.085  0.27803
## Size2       -0.01219    0.28340  -0.043  0.96568
## Size3        1.61457    0.32104   5.029 4.93e-07 ***
## Mode2       -0.20934    0.19033  -1.100  0.27138
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for poisson family taken to be 1)
##
##     Null deviance: 385.53  on 89  degrees of freedom
## Residual deviance: 195.68  on 74  degrees of freedom
```

```
## AIC: 332.02
##
## Number of Fisher Scoring iterations: 7
```

From above, we see that p-values for System 1, 3, 4, Operator 1, 3, 4, Valve 1, 4, 5, Size 1, 3 are smaller than 0.05, we conclude that these factors are asscociated with failures.

2a).Interpret the estimated parameters. For 1(a): When other explanatory variables are fixed, operator 2 will cause e^-0.79 times as many failures as operator 1; operator 3 will cause e^-0.423 times as many failures as operator 1; operator 4 will cause e^-1.74 times as many failures as operator 1.

For 1(b): System: When other explanatory variables are fixed, system 2 will cause e^0.916 times as many failures as system 1; system 3 will cause e^1.02 times as many failures as system 1; system 4 will cause e^1.223 times as many failures as system 1; system 5 will cause e^0.33 times as many failures as system 1.

Operator: When other explanatory variables are fixed, operator 2 will cause e^0.7 times as many failures as operator 1; operator 3 will cause e^-1.19 times as many failures as operator 1; operator 4 will cause e^-2.47 times as many failures as operator 1.

Valve: When other explanatory variables are fixed, valve 2 will cause e^0.1853 times as many failures as valve 1; valve 3 will cause e^0.6067 times as many failures as valve 1; valve 4 will cause e^2.9589 times as many failures as valve 1; valve 5 will cause e^1.79318 times as many failures as valve 1; valve 6 will cause e^1.0089 times as many failures as valve 1.

Size: When other explanatory variables are fixed, size 2 will cause e^-0.012 times as many failures as size 1; size 3 will cause e^1.614 times as many failures as size 1.

Mode: When other explanatory variables are fixed, mode 2 will cause e^-0.2093 times as many failures as mode 1.

2b) Assess the goodness of fit of the model.

```r
anova(poi, test='Chi')
```

```
## Analysis of Deviance Table
##
## Model: poisson, link: log
##
## Response: Failures
##
## Terms added sequentially (first to last)
##
##
##          Df Deviance Resid. Df Resid. Dev  Pr(>Chi)
## NULL                        89     413.76
## Operator  3   27.047        86     386.71 5.755e-06 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```r
anova(poi2, test='Chi')
```

```
## Analysis of Deviance Table
##
## Model: poisson, link: log
##
## Response: Failures
##
## Terms added sequentially (first to last)
##
```

```
##
##          Df Deviance Resid. Df Resid. Dev  Pr(>Chi)
## NULL                        89     385.53
## System    4   22.704        85     362.83 0.0001451 ***
## Operator  3    5.335        82     357.49 0.1488176
## Valve     5  109.857        77     247.63 < 2.2e-16 ***
## Size      2   50.742        75     196.89 9.584e-12 ***
## Mode      1    1.213        74     195.68 0.2708352
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

For the first model, the Operator is significant and associate with failures. For the second model, System, Valve, Size are significant and asscociate with failures.

3.

```
library(glmnet)
```

```
## Loading required package: Matrix
```

```
## Loaded glmnet 3.0-2
```

```
set.seed(123)
X<-model.matrix(Failures~System+Operator+Valve+Size+Mode,data=data)
Y<-data$Failures
cv<-cv.glmnet(X, Y, offset=log(data$Time), family='poisson', data=data)
poi3<-glmnet(X, Y, offset=log(data$Time), lambda=cv$lambda.min, family="poisson")

coef(poi3)
```

```
## 17 x 1 sparse Matrix of class "dgCMatrix"
##                      s0
## (Intercept) -1.4247609
## (Intercept)  .
## System2      .
## System3      .
## System4      .
## System5      .
## Operator2    .
## Operator3    .
## Operator4    .
## Valve2       .
## Valve3       .
## Valve4       0.7307952
## Valve5       .
## Valve6       .
## Size2        .
## Size3        0.5531644
## Mode2        .
```

From above, valve 4 and size 3 are significant. When other explanatory variables are fixed, valve 4 will cause e^0.7307952 times as many failures as valve 1; Size 3 will cause e^0.5531644 times as many failures as size 1.