

multilevel2

kallakuri radhakrishna

17/07/2020

#example 1 Access the data and modelling

```
library(MASS)
```

```
## Warning: package 'MASS' was built under R version 3.6.2
```

```
data <- ships
```

```
#View(data)
```

```
m1 <- glm(incidents~service,family=poisson(link = "log"),data=data)
summary(m1)
```

```
##
```

```
## Call:
```

```
## glm(formula = incidents ~ service, family = poisson(link = "log"),
##      data = data)
```

```
##
```

```
## Deviance Residuals:
```

```
##      Min       1Q   Median       3Q      Max
## -6.0040  -3.1674  -2.0055   0.9155   7.2372
```

```
##
```

```
## Coefficients:
```

```
##              Estimate Std. Error z value Pr(>|z|)
## (Intercept)  1.613e+00  7.150e-02  22.55  <2e-16 ***
## service      6.417e-05  2.870e-06  22.36  <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: 730.25  on 39  degrees of freedom
```

```
## Residual deviance: 374.55  on 38  degrees of freedom
```

```
## AIC: 476.41
```

```
##
```

```
## Number of Fisher Scoring iterations: 6
```

#example 2 Using offset after removing zeroes. An offset term is used for a covariate with known slope. This might arise in situations where you are correcting the number of events for an estimate of population size, the number of CFUs on a plate for the dilution in a dilution assay etc. Most times you will not have an offset term. `offset=log(service)` means we are interested in the rate which gives us the value of -6.13. We see that type D,E and year 75 are insignificant.

```
data1 <-data[!(data$service== 0),]
data1
```

```
##      type year period service incidents
## 1      A   60     60     127         0
## 2      A   60     75      63         0
## 3      A   65     60    1095         3
## 4      A   65     75    1095         4
```

## 5	A	70	60	1512	6
## 6	A	70	75	3353	18
## 8	A	75	75	2244	11
## 9	B	60	60	44882	39
## 10	B	60	75	17176	29
## 11	B	65	60	28609	58
## 12	B	65	75	20370	53
## 13	B	70	60	7064	12
## 14	B	70	75	13099	44
## 16	B	75	75	7117	18
## 17	C	60	60	1179	1
## 18	C	60	75	552	1
## 19	C	65	60	781	0
## 20	C	65	75	676	1
## 21	C	70	60	783	6
## 22	C	70	75	1948	2
## 24	C	75	75	274	1
## 25	D	60	60	251	0
## 26	D	60	75	105	0
## 27	D	65	60	288	0
## 28	D	65	75	192	0
## 29	D	70	60	349	2
## 30	D	70	75	1208	11
## 32	D	75	75	2051	4
## 33	E	60	60	45	0
## 35	E	65	60	789	7
## 36	E	65	75	437	7
## 37	E	70	60	1157	5
## 38	E	70	75	2161	12
## 40	E	75	75	542	1

```
m2 <- glm(incidents~offset(log(service)),family=poisson(link="log"),data=data1)
m2
```

```
##
## Call:  glm(formula = incidents ~ offset(log(service)), family = poisson(link = "log"),
##      data = data1)
##
## Coefficients:
## (Intercept)
##      -6.13
##
## Degrees of Freedom: 33 Total (i.e. Null);  33 Residual
## Null Deviance:      146.3
## Residual Deviance: 146.3    AIC: 246.2
```

```
summary(m2)
```

```
##
## Call:
## glm(formula = incidents ~ offset(log(service)), family = poisson(link = "log"),
##      data = data1)
##
## Deviance Residuals:
##      Min       1Q   Median       3Q      Max
## -6.7636  -0.8573  -0.1795   1.3670   3.9808
```

```
##
## Coefficients:
##           Estimate Std. Error z value Pr(>|z|)
## (Intercept)  -6.130      0.053  -115.7  <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: 146.33  on 33  degrees of freedom
## Residual deviance: 146.33  on 33  degrees of freedom
## AIC: 246.19
##
## Number of Fisher Scoring iterations: 5
data1$period <- as.factor(data1$period)
data1$year <- as.factor(data1$year)
m3 <- glm(incidents~offset(log(service))+type+year+period,family=poisson(link = "log"),data=data1)
summary(m3)

##
## Call:
## glm(formula = incidents ~ offset(log(service)) + type + year +
##      period, family = poisson(link = "log"), data = data1)
##
## Deviance Residuals:
##      Min       1Q   Median       3Q      Max
## -1.6768  -0.8293  -0.4370   0.5058   2.7912
##
## Coefficients:
##           Estimate Std. Error z value Pr(>|z|)
## (Intercept) -6.40590     0.21744 -29.460  < 2e-16 ***
## typeB       -0.54334     0.17759  -3.060  0.00222 **
## typeC       -0.68740     0.32904  -2.089  0.03670 *
## typeD       -0.07596     0.29058  -0.261  0.79377
## typeE        0.32558     0.23588   1.380  0.16750
## year65       0.69714     0.14964   4.659 3.18e-06 ***
## year70       0.81843     0.16977   4.821 1.43e-06 ***
## year75       0.45343     0.23317   1.945  0.05182 .
## period75     0.38447     0.11827   3.251  0.00115 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for poisson family taken to be 1)
##
##      Null deviance: 146.328  on 33  degrees of freedom
## Residual deviance:  38.695  on 25  degrees of freedom
## AIC: 154.56
##
## Number of Fisher Scoring iterations: 5
```

example 3

log(service) is very significant variable. Residual deviance has decreased which is better.

```
m4 <- glm(incidents~log(service)+type+year+period,family=poisson(link ="log"),data=data1)
summary(m4)
```

```
##
## Call:
## glm(formula = incidents ~ log(service) + type + year + period,
##      family = poisson(link = "log"), data = data1)
##
## Deviance Residuals:
##      Min       1Q   Median       3Q      Max
## -1.6580  -0.8939  -0.4900   0.4676   2.7435
##
## Coefficients:
##              Estimate Std. Error z value Pr(>|z|)
## (Intercept)  -5.5940     0.8724  -6.412 1.43e-10 ***
## log(service)   0.9027     0.1018   8.867 < 2e-16 ***
## typeB         -0.3499     0.2702  -1.295  0.19539
## typeC         -0.7631     0.3382  -2.257  0.02404 *
## typeD         -0.1355     0.2971  -0.456  0.64842
## typeE          0.2739     0.2418   1.133  0.25719
## year65         0.6625     0.1536   4.312 1.61e-05 ***
## year70         0.7597     0.1777   4.276 1.90e-05 ***
## year75         0.3697     0.2458   1.504  0.13259
## period75       0.3703     0.1181   3.134  0.00172 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for poisson family taken to be 1)
##
##      Null deviance: 614.539  on 33  degrees of freedom
## Residual deviance:  37.804  on 24  degrees of freedom
## AIC: 155.67
##
## Number of Fisher Scoring iterations: 5

#example 4 Observation 21 has significant residual "type" "year" "period" "service" "incidents" C 70 60 783
6
```

```
a <- m4$residuals
a
```

```
##           1           2           3           4           5           6
## -1.00000000 -1.00000000 -0.24992238 -0.30937157  0.01723337  0.02685994
##           8           9          10          11          12          13
##  0.33182403 -0.06033527  0.14834010  0.08178427 -0.07242729 -0.28215065
##          14          16          17          18          19          20
##  0.04090083  0.09085545 -0.02690284  0.33310815 -1.00000000 -0.42762265
##          21          22          24          25          26          27
##  2.95198775 -0.60044409  0.73333404 -1.00000000 -1.00000000 -1.00000000
##          28          29          30          32          33          35
## -1.00000000  0.45852367  0.80591864 -0.39853871 -1.00000000  0.78898335
##          36          37          38          40
##  1.10589457 -0.17929869 -0.22612301 -0.66804393
```

```
OutVals1 <- boxplot(a, plot=FALSE)$out
```

```
paste("outlier is", OutVals1)
```

```
## [1] "outlier is 2.95198774527149"
```

```
#example 5
```

Of the three factors type is the most significant.

```
res.aov2 <- aov(incidents~log(service)+type+year+period, data =data1)
summary(res.aov2)
```

```
##              Df Sum Sq Mean Sq F value    Pr(>F)
## log(service)  1   5647    5647   93.168 9.77e-10 ***
## type          4    846     211    3.488  0.0221  *
## year          3    166      55    0.913  0.4494
## period        1     57      57    0.941  0.3417
## Residuals     24   1455      61
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
res.aov <- aov(incidents~type+year+period, data =data1)
summary(res.aov)
```

```
##              Df Sum Sq Mean Sq F value    Pr(>F)
## type          4   5901   1475.3   18.951 2.8e-07 ***
## year          3    253     84.5    1.085  0.374
## period        1     70     69.8    0.897  0.353
## Residuals     25   1946     77.8
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```