skrypt.R

Maja Olszanska

```
library(dplyr)
library(tidyr)
library(stringr)
bortkiewicz=read.table("C:/Users/     Maja O   /Documents/R/bortkiewicz.csv", sep="\t")
knitr::kable(head(bortkiewicz))
```

	X1875	X1876	X1877	X1878	X1879	X1880	X1881	X1882	X1883	X1884	X1885	X1886	X1887	X
G	0	2	2	1	0	0	1	1	0	3	0	2	1	
I	0	0	0	2	0	3	0	2	0	0	0	1	1	
II	0	0	0	2	0	2	0	0	1	1	0	0	2	
III	0	0	0	1	1	1	2	0	2	0	0	0	1	
IV	0	1	0	1	1	1	1	0	0	0	0	1	0	
V	0	0	0	0	2	1	0	0	1	0	0	1	0	

```
bortkiewicz$corp=rownames(bortkiewicz)
gather(bortkiewicz, year, value, -corp) %>%mutate(year=as.numeric(str_extract(year, pattern="[0-9]+")))
summary(m1 <-glm(value ~ year + corp, family="poisson", data=bortkiewicz))
```

```
##
## Call:
## glm(formula = value ~ year + corp, family = "poisson", data = bortkiewicz)
##
## Deviance Residuals:
##     Min       1Q   Median       3Q      Max
## -1.6887  -1.1077  -0.8035   0.5348   2.0810
##
## Coefficients:
##               Estimate Std. Error z value Pr(>|z|)
## (Intercept) -3.559e+01  2.343e+01  -1.519   0.1288
## year         1.876e-02  1.243e-02   1.510   0.1312
## corpI        3.850e-09  3.535e-01   0.000   1.0000
## corpII      -2.877e-01  3.819e-01  -0.753   0.4512
## corpIII     -2.877e-01  3.819e-01  -0.753   0.4512
## corpIV      -6.931e-01  4.330e-01  -1.601   0.1094
## corpIX      -2.076e-01  3.734e-01  -0.556   0.5781
## corpV       -3.747e-01  3.917e-01  -0.957   0.3387
## corpVI       6.062e-02  3.483e-01   0.174   0.8618
## corpVII     -2.877e-01  3.819e-01  -0.753   0.4512
## corpVIII    -8.267e-01  4.532e-01  -1.824   0.0681 .
## corpX       -6.454e-02  3.594e-01  -0.180   0.8575
## corpXI       4.463e-01  3.202e-01   1.394   0.1633
## corpXIV      4.055e-01  3.227e-01   1.256   0.2090
## corpXV      -6.931e-01  4.330e-01  -1.601   0.1094
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
```

```
## (Dispersion parameter for poisson family taken to be 1)
##
##     Null deviance: 323.23  on 279  degrees of freedom
## Residual deviance: 294.81  on 265  degrees of freedom
## AIC: 629.89
##
## Number of Fisher Scoring iterations: 5
```

```
1-pchisq(summary(m1)$deviance,summary(m1)$df.residual)
```

```
## [1] 0.1006652
```

```
anova(glm(value ~ 1, family="poisson", data=bortkiewicz), m1)
```

```
## Analysis of Deviance Table
##
## Model 1: value ~ 1
## Model 2: value ~ year + corp
##   Resid. Df Resid. Dev Df Deviance
## 1       279     323.23
## 2       265     294.81 14   28.423
```

```
1-pchisq(28.4, 14)
```

```
## [1] 0.01258477
```

```
step(m1, k =log(nrow(bortkiewicz)))
```

```
## Start:  AIC=684.41
## value ~ year + corp
##
##        Df Deviance    AIC
## - corp 13   320.94 637.29
## - year  1   297.09 681.06
## <none>      294.81 684.41
##
## Step:  AIC=637.29
## value ~ year
##
##        Df Deviance    AIC
## - year  1   323.23 633.94
## <none>      320.94 637.29
##
## Step:  AIC=633.94
## value ~ 1

##
## Call:  glm(formula = value ~ 1, family = "poisson", data = bortkiewicz)
##
## Coefficients:
## (Intercept)
##     -0.3567
##
## Degrees of Freedom: 279 Total (i.e. Null);  279 Residual
## Null Deviance:       323.2
## Residual Deviance: 323.2     AIC: 630.3
```

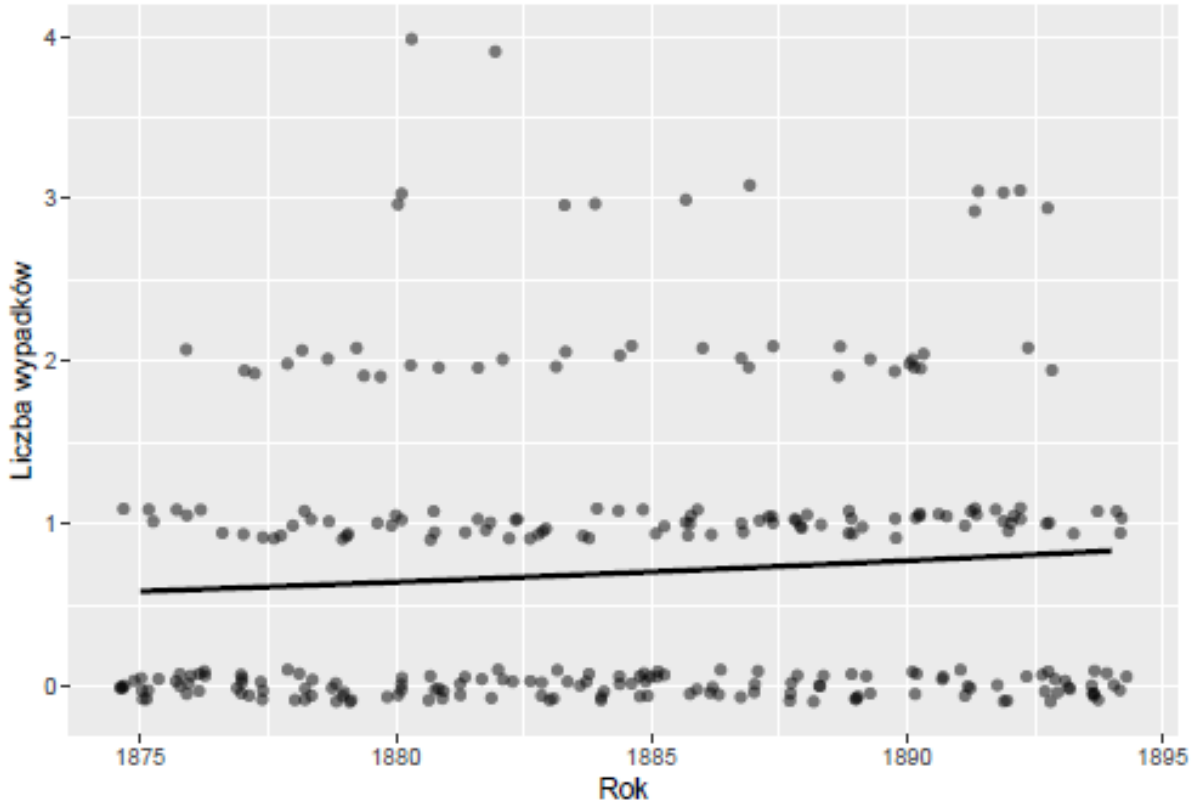
```
m2 <-update(m1, . ~ . - corp)
## test model differences with chi square test
anova(m2, m1, test = "Chisq")
```

```
## Analysis of Deviance Table
##
## Model 1: value ~ year
## Model 2: value ~ year + corp
##   Resid. Df Resid. Dev Df Deviance Pr(>Chi)
## 1       278     320.94
## 2       265     294.81 13   26.137   0.0163 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
1-pchisq(summary(m2)$deviance,summary(m2)$df.residual)
```

```
## [1] 0.038955
```

```
library(ggplot2)
bortkiewicz$valuehat <-predict(m2, type="response")
ggplot(bortkiewicz,aes(x = year, y = valuehat)) +
  geom_point(aes(y = value), alpha=.5, position=position_jitter(h=.1)) +
  geom_line(size = 1) +labs(x = "Rok", y = "Liczba wypadków")
```




```
predict(m1, newdata = bortkiewicz[1:10,])
```

```
##          1          2          3          4          5          6
## -0.4072510 -0.4072510 -0.6949330 -0.6949330 -1.1003981 -0.7819444
```

```
##          7          8          9         10
## -0.3466263 -0.6949330 -1.2339295 -0.6148903
```

```
exp(predict(m1, newdata = bortkiewicz[1:10,]))
```

```
##         1         2         3         4         5         6         7
## 0.6654772 0.6654772 0.4991079 0.4991079 0.3327386 0.4575156 0.7070695
##         8         9        10
## 0.4991079 0.2911463 0.5407002
```

```
predict(m1, newdata = bortkiewicz[1:10,], "response")
```

```
##         1         2         3         4         5         6         7
## 0.6654772 0.6654772 0.4991079 0.4991079 0.3327386 0.4575156 0.7070695
##         8         9        10
## 0.4991079 0.2911463 0.5407002
```