

Lista 10

Matheus Cougias e Klysman Rezende

02/10/2020

#—————QUESTÃO 01————— A análise exploratória do banco de dados já foi realizada na Lista 09, sendo assim, não será abordada nessa lista.

A primeira regressão presente é realizada utilizando a variável Insured como população de referência. Dessa maneira, encontra-se um grande intervalo de Deviance, de -23.951 até 46.914, fazendo com que o modelo gerado não compreenda muito bem os dados de entrada. Nesse modelo, o cliente que possui a maior probabilidade de acionar seu seguro é composto pelas seguintes características: Kilometres na faixa de 1000~15000, residentes da área rural do sul do país, que utiliza o Tipo 9 de veículo, com alto nível de Bonus e com alto número de assegurados em sua região.

Já na regressão em que tomamos o Insured como offset, temos uma diminuição da Deviance, com seu intervalo indo para -8.9411 e 9.4199, identificando que, em relação à primeira regressão realizada, essa compreende melhor os dados. Temos então que o perfil de cliente que maximiza a probabilidade de acionar seu seguro é dado pela quilometragem acima de 25000km, residentes da região de Estolcomo, que utilizam veículos do Tipo 5 e que possuem a menor faixa de Bonus possível (no caso, Bonus = 1). Dessa maneira, pelo fato do intervalo de Deviance ser menor quando utilizamos o Offset, esse modelo é mais recomendável que o inicial.

```
#Leitura e modificação dos dados
dados1 <- read.csv("motorins.dat", sep="")
dados1$Kilometres <- factor(dados1$Kilometres)
levels(dados1$Kilometres) <- c("< 1000", "1000 ~ 15000", "15000 ~ 20000", "20000 ~ 25000", "> 25000")
dados1$Zone <- factor(dados1$Zone)
levels(dados1$Zone) <- c("Stockholm...", "Large cities", "Pequenas cidades do sul", "Area rural do sul",
                        "Area rural do norte", "Gotland")
dados1$Make <- factor(dados1$Make)
levels(dados1$Make) <- c("Tipo 1", "Tipo 2", "Tipo 3", "Tipo 4", "Tipo 5", "Tipo 6", "Tipo 7", "Tipo 8")

#Regressão sem Offset
modelo1 <- glm(Claims ~ Insured + Kilometres + Zone + Make + Bonus, data = dados1, family = poisson)
summary(modelo1)
```

```
##
## Call:
## glm(formula = Claims ~ Insured + Kilometres + Zone + Make + Bonus,
##      family = poisson, data = dados1)
##
## Deviance Residuals:
##      Min       1Q   Median       3Q      Max
## -23.951  -2.837  -1.315   0.045   46.914
##
## Coefficients:
##
##              Estimate Std. Error z value Pr(>|z|)
```

```
## (Intercept)          4.051e+00  1.428e-02 283.765 < 2e-16 ***
## Insured              1.336e-05  1.227e-07 108.914 < 2e-16 ***
## Kilometres1000 ~ 15000  1.878e-01  7.457e-03  25.185 < 2e-16 ***
## Kilometres15000 ~ 20000 -1.933e-01  8.646e-03 -22.357 < 2e-16 ***
## Kilometres20000 ~ 25000 -1.073e+00  1.214e-02 -88.328 < 2e-16 ***
## Kilometres> 25000     -1.217e+00  1.292e-02 -94.159 < 2e-16 ***
## ZoneLarge cities      -1.102e-01  9.496e-03 -11.607 < 2e-16 ***
## ZonePequenas cidades do sul -1.966e-01  9.672e-03 -20.328 < 2e-16 ***
## ZoneArea rural do sul    2.473e-02  9.409e-03   2.628 0.00859 **
## ZonePequenas cidades do norte -1.277e+00  1.454e-02 -87.796 < 2e-16 ***
## ZoneArea rural do norte  -7.804e-01  1.186e-02 -65.791 < 2e-16 ***
## ZoneGotland            -3.502e+00  4.071e-02 -86.038 < 2e-16 ***
## MakeTipo 2             -1.418e+00  2.122e-02 -66.849 < 2e-16 ***
## MakeTipo 3             -1.815e+00  2.505e-02 -72.447 < 2e-16 ***
## MakeTipo 4             -1.703e+00  2.388e-02 -71.335 < 2e-16 ***
## MakeTipo 5             -1.300e+00  2.023e-02 -64.250 < 2e-16 ***
## MakeTipo 6             -8.983e-01  1.733e-02 -51.826 < 2e-16 ***
## MakeTipo 7             -1.649e+00  2.334e-02 -70.666 < 2e-16 ***
## MakeTipo 8             -2.327e+00  3.150e-02 -73.848 < 2e-16 ***
## MakeTipo 9             1.733e+00  1.022e-02 169.539 < 2e-16 ***
## Bonus                 1.242e-01  1.762e-03  70.526 < 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: 435505 on 2181 degrees of freedom
## Residual deviance: 51300 on 2161 degrees of freedom
## AIC: 58980
##
## Number of Fisher Scoring iterations: 5
```

#Regressão com Offset

```
modelo2 <- glm(Claims ~ offset(log(Insured)) + Kilometres + Zone + Make + Bonus, data = dados1, family = poisson)
summary(modelo2)
```

```
##
## Call:
## glm(formula = Claims ~ offset(log(Insured)) + Kilometres + Zone +
## Make + Bonus, family = poisson, data = dados1)
##
## Deviance Residuals:
##      Min       1Q   Median       3Q      Max
## -8.9411  -0.9325  -0.2387   0.6105   9.4199
##
## Coefficients:
##              Estimate Std. Error z value Pr(>|z|)
## (Intercept)    -1.739938    0.013630 -127.651 < 2e-16 ***
## Kilometres1000 ~ 15000    0.204203    0.007511  27.188 < 2e-16 ***
## Kilometres15000 ~ 20000    0.310530    0.008631  35.977 < 2e-16 ***
## Kilometres20000 ~ 25000    0.395459    0.012024  32.889 < 2e-16 ***
## Kilometres> 25000        0.567998    0.012801  44.372 < 2e-16 ***
## ZoneLarge cities    -0.239438    0.009495 -25.216 < 2e-16 ***
## ZonePequenas cidades do sul -0.389831    0.009668 -40.320 < 2e-16 ***
```

```
## ZoneArea rural do sul      -0.585540   0.008650  -67.696 < 2e-16 ***
## ZonePequenas cidades do norte -0.328624   0.014529  -22.618 < 2e-16 ***
## ZoneArea rural do norte    -0.529803   0.011874  -44.620 < 2e-16 ***
## ZoneGotland                -0.735746   0.040697  -18.079 < 2e-16 ***
## MakeTipo 2                 0.078421   0.021239   3.692 0.000222 ***
## MakeTipo 3                -0.240878   0.025092   -9.600 < 2e-16 ***
## MakeTipo 4                -0.656632   0.024181  -27.155 < 2e-16 ***
## MakeTipo 5                 0.158690   0.020234   7.843 4.41e-15 ***
## MakeTipo 6                -0.340402   0.017374  -19.593 < 2e-16 ***
## MakeTipo 7                -0.056424   0.023343   -2.417 0.015642 *
## MakeTipo 8                -0.037993   0.031603   -1.202 0.229286
## MakeTipo 9                -0.068676   0.009955   -6.899 5.25e-12 ***
## Bonus                     -0.198056   0.001288 -153.741 < 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: 34070.6 on 2181 degrees of freedom
## Residual deviance: 4025.1 on 2162 degrees of freedom
## AIC: 11703
##
## Number of Fisher Scoring iterations: 4
```

#—————QUESTÃO 02————— Pelo fato da análise da base de dados já ter sido realizada na Lista 08, não será abordada essa etapa no relatório. Desse modo, com objetivo de comparar o resultado gerado pelas Regressões Binomial e de Poisson, é importante relembrar a conclusão da Lista 08.

Na utilização da Regressão Boniomial, o intervalo de Deviance foi dado entre -2.64532 e 1.85912. A combinação das variáveis que gera o maior número de casos é dada por pessoas da cidade de Fredericia com a faixa de idade entre 70 e 74 anos. Por outro lado, o perfil que gera o menor número de casos são pessoas da cidade de Kolding com a idade entre 40 e 54 anos.

Agora, a análise da atual lista é feita através de uma Regressão de Poisson, onde o intervalo de Deviance gerado foi de -2.63573 e 1.85267. Assim como na regressão anterior, o perfil que gera maior probabilidade de casos de câncer é dado por pessoas de Fredericia com idade entre 70 e 74 anos. Assim como no caso de maior probabilidade, a combinação que gera a menor probabilidade também é de pessoas entre 40 e 54 anos de idade que residem na cidade de Kolding.

Dessa maneira, pode-se perceber que a base foi bem trabalhada tanto pela Regressão Binomial quanto pela Regressão de Poisson, pois em ambos os casos os valores da Deviance são extremamente próximos e pelos valores dos Estimates e P-valores também serem similares. Uma provável causa dessa similaridade nos resultados gerados se dá por ambas as Distribuições Binomial e de Poisson trabalharem sobre dados de probabilidade discreta, onde as tentavidas são independentes e a probabilidade de cada intervalo se mantém constante.

```
#Leitura e modificação dos dados
dados3 <- read.csv("danishlc.dat", sep="\t")
dados3$Age <- factor(dados3$Age,
                    levels = c("40-54", "55-59", "60-64", "65-69", "70-74", ">74"))
dados3$City <- as.factor(dados3$City)

#Regressão de Poisson
modelo3 <- glm(Cases ~ offset(log(Pop)) + Age + City, data = dados3, family = poisson)
summary(modelo2)
```

```
##
## Call:
## glm(formula = Claims ~ offset(log(Insured)) + Kilometres + Zone +
##     Make + Bonus, family = poisson, data = dados1)
##
## Deviance Residuals:
##      Min       1Q   Median       3Q      Max
## -8.9411  -0.9325  -0.2387   0.6105   9.4199
##
## Coefficients:
##              Estimate Std. Error z value Pr(>|z|)
## (Intercept)    -1.739938   0.013630 -127.651 < 2e-16 ***
## Kilometres1000 ~ 15000     0.204203   0.007511  27.188 < 2e-16 ***
## Kilometres15000 ~ 20000    0.310530   0.008631  35.977 < 2e-16 ***
## Kilometres20000 ~ 25000    0.395459   0.012024  32.889 < 2e-16 ***
## Kilometres> 25000         0.567998   0.012801  44.372 < 2e-16 ***
## ZoneLarge cities        -0.239438   0.009495 -25.216 < 2e-16 ***
## ZonePequenas cidades do sul -0.389831   0.009668 -40.320 < 2e-16 ***
## ZoneArea rural do sul     -0.585540   0.008650 -67.696 < 2e-16 ***
## ZonePequenas cidades do norte -0.328624   0.014529 -22.618 < 2e-16 ***
## ZoneArea rural do norte   -0.529803   0.011874 -44.620 < 2e-16 ***
## ZoneGotland              -0.735746   0.040697 -18.079 < 2e-16 ***
## MakeTipo 2                0.078421   0.021239   3.692 0.000222 ***
## MakeTipo 3               -0.240878   0.025092  -9.600 < 2e-16 ***
## MakeTipo 4               -0.656632   0.024181 -27.155 < 2e-16 ***
## MakeTipo 5                0.158690   0.020234   7.843 4.41e-15 ***
## MakeTipo 6               -0.340402   0.017374 -19.593 < 2e-16 ***
## MakeTipo 7               -0.056424   0.023343  -2.417 0.015642 *
## MakeTipo 8               -0.037993   0.031603  -1.202 0.229286
## MakeTipo 9               -0.068676   0.009955  -6.899 5.25e-12 ***
## Bonus                  -0.198056   0.001288 -153.741 < 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: 34070.6  on 2181  degrees of freedom
## Residual deviance:  4025.1  on 2162  degrees of freedom
## AIC: 11703
##
## Number of Fisher Scoring iterations: 4
```