ITHIM_PhaseII_Model1

Ofurhe Igbinedion

May 30, 2018

```
setwd("C:/Users/ofurhe/Desktop/ITHIM")

#read regression Data
RegressionData1 <- read.csv("regression_data_model_1.csv", header = TRUE)
head(RegressionData1)
```

```
##   X  zcta total.population Death.Count non.Hisp.White non.Hisp.Black
## 1 1 90001            57110    228.0000            356           5857
## 2 2 90002            51223    267.3333            330          12723
## 3 3 90003            66266    331.3333            462          15672
## 4 4 90004            62180    250.3333          10466           1962
## 5 5 90005            37681    149.6667           3039           1788
## 6 6 90006            59185    229.3333           1861           1862
##   Hisp.or.Latino Poverty  Edu1 male1 male2 male3 male4 male5 male6 male7
## 1          50544   15327 17816  2922  5411  7708  6188  4151  1263   556
## 2          37598   15046 11573  2748  5003  6909  4787  3662  1026   489
## 3          49386   24974 18076  3469  6426  8983  6775  4760  1358   614
## 4          31987   14670 12480  2052  3565  7825  8455  5930  2097   911
## 5          19578   10197  9151  1215  2084  5231  5328  3388  1099   657
## 6          44074   18083 17752  2298  3865  8012  7692  5253  1897   869
##   male8 female1 female2 female3 female4 female5 female6 female7 female8
## 1   269    2811    5401    7412    6142    4167    1428     747     534
## 2   252    2570    4964    7061    5181    4085    1251     741     494
## 3   246    3368    6189    8822    7091    5257    1522     851     535
## 4   467    1920    3440    6979    7468    6388    2580    1286     817
## 5   297    1124    1916    4351    4441    3490    1427    1011     622
## 6   368    2166    3798    6917    6567    5395    2204    1127     757
```

```
# create new data frame with all missing values removed
RegressionData1.c <- na.omit(RegressionData1)

######## Regression in a single equation

#Caclulate Rates
RegressionData1.c$Death.Rate <- (RegressionData1.c$Death.Count / RegressionData1.c$total.populat
ion)
RegressionData1.c$White.Rate <- (RegressionData1.c$non.Hisp.White / RegressionData1.c$total.popu
lation)
RegressionData1.c$Black.Rate <- (RegressionData1.c$non.Hisp.Black / RegressionData1.c$total.popu
lation)
RegressionData1.c$Hisp.Rate <- (RegressionData1.c$Hisp.or.Latino / RegressionData1.c$total.popul
ation)
RegressionData1.c$Poverty.Rate <- (RegressionData1.c$Poverty / RegressionData1.c$total.populatio
n)
RegressionData1.c$Education.Rate <- (RegressionData1.c$Edu1 / RegressionData1.c$total.population
)

RegressionData1.c$male12.Rate <- ((RegressionData1.c$male1 + RegressionData1.c$male2)/ Regressio
nData1.c$total.population)
RegressionData1.c$male34.Rate <- ((RegressionData1.c$male3 + RegressionData1.c$male4)/ Regressio
nData1.c$total.population)
RegressionData1.c$male56.Rate <- ((RegressionData1.c$male5 + RegressionData1.c$male6)/ Regressio
nData1.c$total.population)
RegressionData1.c$male78.Rate <- ((RegressionData1.c$male7 + RegressionData1.c$male8)/ Regressio
nData1.c$total.population)

RegressionData1.c$female12.Rate <- ((RegressionData1.c$female1 + RegressionData1.c$female2)/ Reg
ressionData1.c$total.population)
RegressionData1.c$female34.Rate <- ((RegressionData1.c$female3 + RegressionData1.c$female4)/ Reg
ressionData1.c$total.population)
RegressionData1.c$female56.Rate <- ((RegressionData1.c$female5 + RegressionData1.c$female6)/ Reg
ressionData1.c$total.population)
RegressionData1.c$female78.Rate <- ((RegressionData1.c$female7 + RegressionData1.c$female8)/ Reg
ressionData1.c$total.population)

# select variables
data.v <- cbind(RegressionData1.c$zcta, RegressionData1.c$Death.Rate, RegressionData1.c$White.Ra
te, RegressionData1.c$Black.Rate, RegressionData1.c$Hisp.Rate, RegressionData1.c$Poverty.Rate, R
egressionData1.c$Education.Rate, RegressionData1.c$male12.Rate, RegressionData1.c$male34.Rate, R
egressionData1.c$male56.Rate, RegressionData1.c$male78.Rate, RegressionData1.c$female12.Rate, Re
gressionData1.c$female34.Rate, RegressionData1.c$female56.Rate, RegressionData1.c$female78.Rate
)

colnames(data.v) <- c("Zip","Death.Rate","White.Rate","Black.Rate","Hisp.Rate","Poverty.Rate","E
ducation.Rate", "male12.Rate", "male34.Rate", "male56.Rate","male78.Rate", "female12.Rate", "fem
ale34.Rate", "female56.Rate","female78.Rate")
head(data.v)
```

```
##        Zip  Death.Rate  White.Rate Black.Rate Hisp.Rate Poverty.Rate
## [1,] 90001 0.003992296 0.006233584 0.10255647 0.8850289    0.2683768
## [2,] 90002 0.005219010 0.006442418 0.24838451 0.7340062    0.2937352
## [3,] 90003 0.005000050 0.006971901 0.23650137 0.7452691    0.3768750
## [4,] 90004 0.004025946 0.168317787 0.03155355 0.5144259    0.2359280
## [5,] 90005 0.003971940 0.080650726 0.04745097 0.5195722    0.2706138
## [6,] 90006 0.003874856 0.031443778 0.03146067 0.7446819    0.3055335
##      Education.Rate male12.Rate male34.Rate male56.Rate male78.Rate
## [1,]      0.3119594  0.14591140   0.2433199  0.09479951  0.01444581
## [2,]      0.2259337  0.15131874   0.2283349  0.09152139  0.01446616
## [3,]      0.2727794  0.14932243   0.2377992  0.09232487  0.01297800
## [4,]      0.2007076  0.09033451   0.2618205  0.12909296  0.02216147
## [5,]      0.2428545  0.08755076   0.2802208  0.11907858  0.02531780
## [6,]      0.2999409  0.10413111   0.2653375  0.12080764  0.02090057
##      female12.Rate female34.Rate female56.Rate female78.Rate
## [1,]    0.14379268     0.2373315    0.09796883    0.02243040
## [2,]    0.14708237     0.2389942    0.10417195    0.02411026
## [3,]    0.14422177     0.2401382    0.10229982    0.02091570
## [4,]    0.08620135     0.2323416    0.14422644    0.03382116
## [5,]    0.08067726     0.2333271    0.13049017    0.04333749
## [6,]    0.10076878     0.2278280    0.12839402    0.03183239
```

```
data.v <- data.v[is.finite(rowSums(data.v)),]
data.v <- as.data.frame(data.v)


# Define full and null models and do step procedure
model.null <- glm(Death.Rate~1,
                  data = data.v,
                  family = binomial(link = "logit")
                  )
```

```
## Warning in eval(family$initialize): non-integer #successes in a binomial
## glm!
```

```
summary(model.null)
```

```
##
## Call:
## glm(formula = Death.Rate ~ 1, family = binomial(link = "logit"),
##     data = data.v)
##
## Deviance Residuals:
##      Min        1Q    Median        3Q       Max
## -0.13102  -0.04257  -0.02115   0.00351   2.25440
##
## Coefficients:
##             Estimate Std. Error z value Pr(>|z|)
## (Intercept)  -4.7536     0.2595  -18.32   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 14.6  on 1752  degrees of freedom
## Residual deviance: 14.6  on 1752  degrees of freedom
## AIC: 41.601
##
## Number of Fisher Scoring iterations: 7
```

```
model.full <- glm(Death.Rate ~ Black.Rate + Hisp.Rate + Poverty.Rate + Education.Rate + male12.R
ate + male34.Rate
                  + male56.Rate + male78.Rate + female12.Rate + female34.Rate + female56.Rate +
female78.Rate,
              data = data.v,
              family = binomial ("logit")
              )
```

```
## Warning in eval(family$initialize): non-integer #successes in a binomial
## glm!
```

```
summary(model.full)
```

```
##
## Call:
## glm(formula = Death.Rate ~ Black.Rate + Hisp.Rate + Poverty.Rate +
##     Education.Rate + male12.Rate + male34.Rate + male56.Rate +
##     male78.Rate + female12.Rate + female34.Rate + female56.Rate +
##     female78.Rate, family = binomial("logit"), data = data.v)
##
## Deviance Residuals:
##      Min        1Q    Median        3Q       Max
## -0.25969  -0.02863  -0.00887   0.00756   1.95208
##
## Coefficients: (1 not defined because of singularities)
##                Estimate Std. Error z value Pr(>|z|)
## (Intercept)    -0.77754    2.07920  -0.374    0.708
## Black.Rate      0.06482    4.05273   0.016    0.987
## Hisp.Rate      -0.68573    2.09598  -0.327    0.744
## Poverty.Rate   -0.30577    2.87526  -0.106    0.915
## Education.Rate -0.31946    3.57296  -0.089    0.929
## male12.Rate    -2.90558   18.41570  -0.158    0.875
## male34.Rate    -3.25216    3.95035  -0.823    0.410
## male56.Rate    -2.72729    5.04939  -0.540    0.589
## male78.Rate     1.22682    8.90225   0.138    0.890
## female12.Rate  -2.58035   19.28865  -0.134    0.894
## female34.Rate  -6.62348    6.56433  -1.009    0.313
## female56.Rate  -6.01627    6.56608  -0.916    0.360
## female78.Rate        NA         NA      NA       NA
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 14.6000  on 1752  degrees of freedom
## Residual deviance:  9.8954  on 1741  degrees of freedom
## AIC: 61.815
##
## Number of Fisher Scoring iterations: 8
```

```
step(model.null,
     scope = list(upper=model.full),
     direction="both",
     test="Chisq",
     data=data.v)
```

```
## Start:  AIC=41.6
## Death.Rate ~ 1
```

```
## Warning in eval(family$initialize): non-integer #successes in a binomial
## glm!

## Warning in eval(family$initialize): non-integer #successes in a binomial
## glm!

## Warning in eval(family$initialize): non-integer #successes in a binomial
## glm!

## Warning in eval(family$initialize): non-integer #successes in a binomial
## glm!

## Warning in eval(family$initialize): non-integer #successes in a binomial
## glm!

## Warning in eval(family$initialize): non-integer #successes in a binomial
## glm!

## Warning in eval(family$initialize): non-integer #successes in a binomial
## glm!

## Warning in eval(family$initialize): non-integer #successes in a binomial
## glm!

## Warning in eval(family$initialize): non-integer #successes in a binomial
## glm!

## Warning in eval(family$initialize): non-integer #successes in a binomial
## glm!

## Warning in eval(family$initialize): non-integer #successes in a binomial
## glm!

## Warning in eval(family$initialize): non-integer #successes in a binomial
## glm!

## Warning in eval(family$initialize): non-integer #successes in a binomial
## glm!
```

```
##                  Df Deviance    AIC    LRT Pr(>Chi)
## + female78.Rate   1   11.435 40.435 3.1655  0.07521 .
## + female34.Rate   1   12.096 41.097 2.5039  0.11356
## <none>                14.600 41.601
## + male78.Rate     1   12.668 41.669 1.9319  0.16455
## + female12.Rate   1   13.319 42.320 1.2809  0.25774
## + male12.Rate     1   13.333 42.333 1.2674  0.26025
## + Hisp.Rate       1   13.341 42.342 1.2591  0.26183
## + male34.Rate     1   13.511 42.512 1.0886  0.29679
## + Education.Rate  1   14.154 43.155 0.4461  0.50419
## + male56.Rate     1   14.208 43.208 0.3923  0.53110
## + Poverty.Rate    1   14.249 43.250 0.3507  0.55372
## + female56.Rate   1   14.413 43.414 0.1867  0.66571
## + Black.Rate      1   14.534 43.535 0.0655  0.79798
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
## Warning in eval(family$initialize): non-integer #successes in a binomial
## glm!
```

```
##
## Step:  AIC=43.23
## Death.Rate ~ female78.Rate
```

```
##
## Call:  glm(formula = Death.Rate ~ female78.Rate, family = binomial(link = "logit"),
##     data = data.v)
##
## Coefficients:
##   (Intercept)  female78.Rate
##        -5.031          4.699
##
## Degrees of Freedom: 1752 Total (i.e. Null);  1751 Residual
## Null Deviance:       14.6
## Residual Deviance: 11.43     AIC: 43.23
```

```
model.final <- glm(Death.Rate ~ female78.Rate,
                   data = data.v,
                   family = binomial(link = "logit")
                   )
```

```
## Warning in eval(family$initialize): non-integer #successes in a binomial
## glm!
```

```
summary(model.final)
```

```
##
## Call:
## glm(formula = Death.Rate ~ female78.Rate, family = binomial(link = "logit"),
##     data = data.v)
##
## Deviance Residuals:
##      Min        1Q    Median        3Q       Max
## -0.17342  -0.03471  -0.01804   0.00133   2.16225
##
## Coefficients:
##               Estimate Std. Error z value Pr(>|z|)
## (Intercept)    -5.0312     0.2992 -16.816   <2e-16 ***
## female78.Rate   4.6995     2.0196   2.327     0.02 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 14.600  on 1752  degrees of freedom
## Residual deviance: 11.435  on 1751  degrees of freedom
## AIC: 43.228
##
## Number of Fisher Scoring iterations: 7
```