BDA - Project Work

Jacopo Losi, Nicola Saljoughi

Contents


```
mydata <- data
mydata$Season <- data$Month
mydata$Month = NULL


# Hospitalised

indexHosp   <- which(data$Hospitalised == 'yes')
indexNoHosp <- which(data$Hospitalised == 'no')

mydata$Hospitalised[indexHosp]   <- 1    # 1 --> yes
mydata$Hospitalised[indexNoHosp] <- 0    # 0 --> no


# Died

indexDied   <- which(data$Died == 'yes')
indexNoDied <- which(data$Died == 'no')

mydata$Died[indexDied]   <- 1    # 1 --> yes
mydata$Died[indexNoDied] <- 0    # 0 --> no


# Urban

indexUrban   <- which(data$Urban == 'yes')
indexNoUrban <- which(data$Urban == 'no')

mydata$Urban[indexUrban]   <- 1    # 1 --> yes
mydata$Urban[indexNoUrban] <- 0    # 0 --> no


# Season

indexSpring <- which(data$Month >= 3 & data$Month <= 5)
indexSummer <- which(data$Month >= 6 & data$Month <= 8)
indexAutumn <- which(data$Month >= 9 & data$Month <= 11)
indexWinter <- which(data$Month == 12 | data$Month <= 2)
```

```r
mydata$Season[indexSpring] <- 1  # 1 --> Spring
mydata$Season[indexSummer] <- 2  # 2 --> Summer
mydata$Season[indexAutumn] <- 3  # 3 --> Autumn
mydata$Season[indexWinter] <- 4  # 4 --> Winter


# Sex

indexMale   <- which(data$Sex == 'male')
indexFemale <- which(data$Sex == 'female')

mydata$Sex[indexMale]   <- 1    # 1 --> male
mydata$Sex[indexFemale] <- 0    # 0 --> female


# Age

indexAgeOne   <- which(data$Age <= 34)
indexAgeTwo   <- which(data$Age >= 35 & data$Age <= 49)
indexAgeThree <- which(data$Age >= 50 & data$Age <= 64)
indexAgeFour  <- which(data$Age >= 65)

mydata$Age[indexAgeOne]   <- 1   # 1 --> <34
mydata$Age[indexAgeTwo]   <- 2   # 2 --> 35-49
mydata$Age[indexAgeThree] <- 3   # 3 --> 50-64
mydata$Age[indexAgeFour]  <- 4   # 4 --> >65


# Education

indexEduZero  <- which(data$Education == 'iliterate')
indexEduOne   <- which(data$Education == 'primary')
indexEduTwo   <- which(data$Education == 'Secondary')
indexEduThree <- which(data$Education == 'Tertiary')

mydata$Education[indexEduZero]   <- 0   # 0 --> iliterate
mydata$Education[indexEduOne]   <- 1   # 1 --> primary
mydata$Education[indexEduTwo] <- 2   # 2 --> Secondary
mydata$Education[indexEduThree]  <- 3   # 3 --> Tertiary


# Occupation

indexFarm   <- which(data$Occupation == 'farming')
indexNoFarm <- which(data$Occupation != 'farming')

mydata$Occupation[indexFarm]   <- 1    # 1 --> farming
mydata$Occupation[indexNoFarm] <- 0    # 0 --> non farming


# Method

indexPesticide <- which(data$method == 'Pesticide')
```

```r
indexPoison    <- which(data$method == 'Other poison')
indexHanging    <- which(data$method == 'hanging')
indexOthers    <- which(data$method != 'Pesticide' &
                        data$method != 'Other poison' &
                        data$method != 'hanging')

mydata$method[indexPesticide] <- 1 # 1 --> Pesticide
mydata$method[indexPoison]    <- 2 # 2 --> Other poison
mydata$method[indexHanging]   <- 3 # 3 --> hanging
mydata$method[indexOthers]    <- 4 # 4 --> All others


str(mydata)
```

```
## 'data.frame':    2571 obs. of  11 variables:
##  $ Person_ID   : int  1 2 3 4 5 6 7 8 9 10 ...
##  $ Hospitalised: chr  "1" "0" "0" "0" ...
##  $ Died        : chr  "0" "1" "1" "1" ...
##  $ Urban       : chr  "0" "0" "0" "0" ...
##  $ Year        : int  2010 2009 2010 2011 2009 2009 2010 2010 2010 2011 ...
##  $ Sex         : chr  "0" "1" "1" "1" ...
##  $ Age         : num  2 4 3 4 3 3 4 3 4 1 ...
##  $ Education   : chr  "2" "1" "1" "1" ...
##  $ Occupation  : chr  "0" "1" "1" "1" ...
##  $ method      : chr  "2" "3" "3" "3" ...
##  $ Season      : num  4 1 4 4 2 3 4 3 2 4 ...
```

Introduction

Objective

The objective of the study is to estimate the incidence of seriours suicide attempts (SSAs), defined as suicide attempts resulting in either death or hospitalization, and to analyse the factors associated with fatality among the attempters.

Data

The data set is constituted by 2571 observations of 11 variables:

- `Person_ID`: ID number, $1, ..., 2571$
- `Hospitalised`: *yes* or *no*
- `Died`: *yes* or *no*
- `Urban`: *yes*, *no* or *unknown*
- `Year`: 2009, 2010 or 2011
- `Month`: $1, ..., 12$
- `Sex`: *female* or *male*
- `Age`: years
- `Education`: *iliterate*, *primary*, *Secondary*, *Tertiary* or *unknown*
- `Occupation`: one of ten categories

- `method`: one of nine methods

Source

Sun J, Guo X, Zhang J, Wang M, Jia C, Xu A (2015) "Incidence and fatality of serious suicide attempts in a predominantly rural population in Shandong, China: a public health surveillance study," BMJ Open 5(2): e006762. https://doi.org/10.1136/bmjopen-2014-006762

Data downloaded via Dryad Digital Repository. https://doi.org/10.5061/dryad.r0v35

Analysis

```
rural_men   <- subset(data, data$Sex=="male" & data$Urban=="no")
rural_women <- subset(data, data$Sex=="female" & data$Urban=="no")
urban_men   <- subset(data, data$Sex=="male" & data$Urban=="yes")
urban_women <- subset(data, data$Sex=="female" & data$Urban=="yes")
```

```
str(data)
```

```
## 'data.frame':    2571 obs. of  11 variables:
##  $ Person_ID   : int  1 2 3 4 5 6 7 8 9 10 ...
##  $ Hospitalised: chr  "yes" "no" "no" "no" ...
##  $ Died        : chr  "no" "yes" "yes" "yes" ...
##  $ Urban       : chr  "no" "no" "no" "no" ...
##  $ Year        : int  2010 2009 2010 2011 2009 2009 2010 2010 2010 2011 ...
##  $ Month       : int  12 3 2 1 8 11 1 10 7 1 ...
##  $ Sex         : chr  "female" "male" "male" "male" ...
##  $ Age         : int  39 83 60 73 51 62 90 54 66 30 ...
##  $ Education   : chr  "Secondary" "primary" "primary" "primary" ...
##  $ Occupation  : chr  "household" "farming" "farming" "farming" ...
##  $ method      : chr  "Other poison" "hanging" "hanging" "hanging" ...
```

```
## Create Stan data
dat <- list(N        = nrow(data),
            p        = 10,
            died     = data$Died,
            urban    = data$Urban,
            year     = data$Year,
            month    = data$Season,
            sex      = data$Sex,
            age      = data$Age,
            edu      = data$Education,
            job      = data$Occupation,
            method   = data$method)

## Load Stan file
fileName <- "./logistic_regression_model.stan"
stan_code <- readChar(fileName, file.info(fileName)$size)
cat(stan_code)
```

```
## // Logistic Regression Model
##
## data {
##   // Define variables in data
##   // Number of observations (an integer)
```

```
##   int<lower=0> N;
##   // Number of parameters
##   int<lower=0> p;
##   // Variables
##   int died[N];
##   int<lower=0>  urban[N];
##   int<lower=0>  year[N];
##   int<lower=0> season[N];
##   int<lower=0>  sex[N];
##   int<lower=0>  age[N];
##   int<lower=0>  edu[N];
##   int<lower=0>  job[N];
##   int<lower=0>  method[N];
## }
##
## parameters {
##   // Define parameters to estimate
##   real beta[p];
## }
##
## transformed parameters  {
##   // Probability trasformation from linear predictor
##   real<lower=0> odds[N];
##   real<lower=0, upper=1> prob[N];
##
##   for (i in 1:N) {
##     odds[i] <- exp(beta[1] + beta[2]*urban[i]  + beta[3]*year[i] +
##                              beta[4]*season[i] + beta[5]*sex[i]  +
##                              beta[6]*age[i]    + beta[6]*edu[i]  +
##                              beta[7]*job[i]    + beta[8]*method[i] );
##     prob[i] <- odds[i] / (odds[i] + 1);
##   }
## }
##
## model {
##   // Prior part of Bayesian inference (flat if unspecified)
##
##   // Likelihood part of Bayesian inference
##     died ~ bernoulli(prob);
## }
```

	All SAAs	**Hospitalised and survived**	**Hospitalised but died**	**Total SSA hospitalisations**	**SSA deaths without hospitalisation**	**Total SSA deaths**
Urban						
Female	149	99	18	117	32	50
Male	128	65	17	82	46	63
Both	277	164	35	199	78	113
Rural						
Female	1134	598	100	698	436	536
Male	1079	474	103	577	502	605
Both	2213	1072	203	1275	938	1141
Total						
Female	1328	741	118	859	469	587
Male	1243	574	120	694	549	669
Both	2571	1315	238	1553	1018	1256

Conclusions