Reporte Técnico

José Luis Palomino 124989, Aurora Peral Rodriguez 125518

Monday, November 03, 2014

# "Análisis por intoxicación de picadura de alacrán"

# Introducción

### Contexto

De acuerdo con la información obtenida de la Secretaria de Salud <http://www.epidemiologia.salud.gob.mx/anuario/html/anuarios.html>. Se busca limpiar, organizar, analizar y dar un reporte sobre las intoxicaciones por picadura de alacrán en los años 2003 a 2013. Para esto usaremos RStudio como la herramienta estadística de análisis.

### RStudio

La Herramienta de RStudio es un programa que sirve para el análisis estadístico, gratuito y que va creciendo con las actualizaciones de los usuarios. Contiende una serie de paquetes especializados y que nos aydudan en tareas como la lectura y procesamiento de datos, gráficas, compraciones, etc...

### Objetivo:

Dar un análisis claro para conocer el fenómeno y tomar medidas preventivas a nivel nacional.

A continuación se expone lo que se tuvo que hacer para la extracción, limpieza, organización, análisis y conclusión del trabajo.

# Definición del problema

Los principales aspectos que deseamos obtener con este anális son:

°¿En qué estados hay mayor intoxicación? °¿En qué meses se presenta un mayor número de casos?

Consideramos que éstas dos preguntas son las más relevantes porque responden a la necesidad de saber en qué mes y en dónde se necesita una mayor cantidad de antídoto.

# Obtención y descripción de los datos

Los datos fueron obtenidos de la página <http://www.epidemiologia.salud.gob.mx/anuario/html/anuarios.html>. Estos datos se presentan en formato pdf y fue necesario convertirlos a un archivo separado por comas (.csv) en una pltaforma Bash de Linux que se encuentra en la siguiente carpeta "C:/Users/HP/Desktop/EA1/BRUTOS/(EDAD,FUENTE,MES)/procesa"

Determinamos consolidar la información en un total de tres archivos; edad, mes y fuente de información.

Además lo convertimos a formato largo con la función "gather" con el objeto de poder graficar de forma que sirva al análisis

# Transformación de los datos

El procesamiento de datos se hizo mediante el código:

### DATOS PROCESADOS POR GRUPOS DE EDAD

## Definimos el cache

directorio <- "C:/Users/Win7/Desktop/EA1/DATOSBRUTOS/EDAD" setwd(directorio) knitr::opts\_knit$set(root.dir = directorio) # Opciones del documento. knitr::opts\_chunk$set(fig.path = "FIGURAS/Figuras\_", fig.align = 'center', fig.width = 12, fig.height = 8, cache.path = "CACHE/Cache\_", cache = TRUE, eval = TRUE, warning = FALSE, message = FALSE, echo = TRUE, comment = NA)

# Hacemos la organización de los datos

sexo<- c("F","M","G") aÃ±o<- c("2003", "2004", "2005", "2006", "2007", "2008", "2009", "2010", "2011","2012","2013") name <- " " Datos.Edad <- data.frame() for (j in 1:3) { for(i in 1:11) { name <- sprintf("grupoedad%s(%s).pdf.txt.csv", sexo[j], aÃ±o[i]) print(name) tablaEdad <- read.csv(name, check.names=FALSE) tail(tablaEdad) tablaEdadsexo = as.factor(sexo[j]) str(tablaEdad) sapply(tablaEdad,class) auxEdad <- Datos.Edad Datos.Edad <- gather(data = tablaEdad, key = "cortesedad", value = "numerocasos", 2:14) Datos.Edad<- rbind(auxEdad, Datos.Edad) } }

## Guardamos la tabla ya limpia y organizada en formato largo

write.table(Datos.Edad, file = "C:/Users/HP/Desktop/EA1/PROCESADOS/Edad.csv", sep = ",", row.names = FALSE)

## Leemos unicamente los datos de los Estados (omitiendo los de "TOTAL GLOBAL")

Datos.Edad.Final <-subset(Datos.Edad, Estado != "TOTAL GLOBAL") Datos.Edad.Final

## DATOS PROCESADOS POR FUENTE DE NOTIFICACION

directorio <- "C:/Users/HP/Desktop/EA1/BRUTOS/FUENTE" setwd(directorio) knitr::opts\_knit$set(root.dir = directorio) # Opciones del documento. knitr::opts\_chunk$set(fig.path = "FIGURAS/Figuras\_", fig.align = 'center', fig.width = 12, fig.height = 8, cache.path = "CACHE/Cache\_", cache = TRUE, eval = TRUE, warning = FALSE, message = FALSE, echo = TRUE, comment = NA)

# Hacemos la organización de los datos

sexo<- c("F","M","G") aÃ±o<- c("2003", "2004", "2005", "2006", "2007", "2008", "2009", "2010", "2011","2012","2013")

Datos.Fuente<-data.frame() for (j in 1:3){

for(i in 1:11){ name <- sprintf("fuente%s(%s).pdf.txt.csv", sexo[j], aÃ±o[i]) tablaFuente <- read.csv(name,check.names=FALSE) tail(tablaFuente) tablaFuentesexo=as.factor(sexo[j]) str(tablaFuente) sapply(tablaFuente,class) auxFuente<-Datos.Fuente Datos.Fuente <- gather(data = tablaFuente, key = "InstituciÃ³n", value = "numerocasos", 2:10) Datos.Fuente<- rbind(auxFuente,Datos.Fuente) }} Datos.Fuente

## Guardamos la tabla ya limpia y organizada en formato largo

write.table(Datos.Fuente, file = "C:/Users/HP/Desktop/EA1/PROCESADOS/Fuente.csv", sep = ",", row.names = FALSE)

### DATOS PROCESADOS POR MES

## Definimos el cache

directorio <- "C:/Users/HP/Desktop/EA1/BRUTOS/MES" setwd(directorio) knitr::opts\_knit$set(root.dir = directorio) # Opciones del documento. knitr::opts\_chunk$set(fig.path = "FIGURAS/Figuras\_", fig.align = 'center', fig.width = 12, fig.height = 8, cache.path = "CACHE/Cache\_", cache = TRUE, eval = TRUE, warning = FALSE, message = FALSE, echo = TRUE, comment = NA)

# Hacemos la organización de los datos

sexo<- c("F","M","G") aÃ±o<- c("2003", "2004", "2005", "2006", "2007", "2008", "2009", "2010", "2011","2012","2013") name<- "" Datos.Mes<-data.frame() for (j in 1:3){ for(i in 1:11){ name <- sprintf("CASOSPORMES%s(%s).pdf.txt.csv", sexo[j], aÃ±o[i])

tablaMes <- read.csv(name,check.names=FALSE) tail(tablaMes) tablaMessexo=as.factor(sexo[j]) str(tablaMes) sapply(tablaMes,class) auxMes<-Datos.Mes Datos.Mes <- gather(data = tablaMes, key = "Mes", value = "numerocasos", 4:15) Datos.Mes<- rbind(auxMes,Datos.Mes)

}} Datos.Mes

## Guardamos la tabla ya limpia y organizada en formato largo

write.table(Datos.Mes, file = "C:/Users/HP/Desktop/EA1/PROCESADOS/mes.csv", sep = ",", row.names = FALSE)

# Analisis Exploratorio

Este es el codigo de la grafica suavizada (se uso la funcion log ya que hay valores muy dispares)

"Hacemos el analisis de los datos procesados por cortes de edad"

## Hacemos un analisis sobre el comportamiento de los casos de picadura de alacran en los diferentes estados SUAVIZADA CON LA FUNCION (log)

install.packages("dplyr")

directorio <- "C:/Users/HP/Desktop/EA1/PROCESADOS" setwd(directorio) GRUPOSEDAD <- read.csv("Edad.csv", header = TRUE, encoding = "utf-8") str(GRUPOSEDAD)

GRUPOSEDAD <- subset(GRUPOSEDAD, Estado!= "TOTAL GLOBAL") GRUPOSEDAD <- subset(GRUPOSEDAD, sexo != "G" ) GRUPOSEDAD <- subset(GRUPOSEDAD, cortesedad != "Total")

str(GRUPOSEDAD)

entidades <- c('Aguascalientes', 'Baja California', 'Baja California Sur', 'Campeche', 'Chiapas', 'Chihuahua', 'Coahuila', 'Colima', 'Distrito Federal', 'Durango', 'Guanajuato', 'Guerrero', 'Hidalgo', 'Jalisco', 'MÃ©xico', 'MichoacÃ¡n', 'Morelos', 'Nayarit', 'Nuevo LeÃ³n', 'Oaxaca', 'Puebla', 'QuerÃ©taro', 'Quintana Roo', 'San Luis PotosÃ????', 'Sinaloa', 'Sonora', 'Tabasco', 'Tamaulipas', 'Tlaxcala', 'Veracruz', 'YucatÃ¡n', 'Zacatecas')

entidades

par(mfrow = c(4, 8), mar = c(0, 0, 1, 0), oma = c(0, 0, 1.2, 0))

anio <- 2013 GRUPOSEDAD.año <- subset(GRUPOSEDAD, año == anio) for(i in 1:32) { GRUPOSEDAD.entidad <- subset(GRUPOSEDAD.año, Estado == entidades[i])

with(GRUPOSEDAD.entidad, plot(0, 0, type = "n", xlim = c(-10, 10), ylim = c(0, 13), xlab ="", ylab = "", axes = FALSE, bty="n", cex.main = 0.5, main = entidades[i]) )

rect(0, 0:11, -log(GRUPOSEDAD.entidad[1:12, ]$numerocasos), 1:12, col="blue", border="white") rect(0, 0:11, log(GRUPOSEDAD.entidad[13:24, ]$numerocasos), 1:12, col="red", border="white") } # Incluimos el año en el márgen superior externo. mtext(paste0(anio, " (Hombres, Mujeres)"), side = 3, outer = TRUE)

la segunda grafica no esta suavizada y pusimos como limite -1000 y 1000 xq eisten edos donde el numero de casos es muy elevado. Se ve rara por que otros estados tienen menos:

## Hacemos el analisis de los datos procesados por cortes de edad

## Hacemos un analisis sobre el comportamiento de los casos de picadura de alacran en los diferentes estados

directorio <- "C:/Users/Win7/Desktop/EA1/DATOSPROCESADOS"

setwd(directorio) GRUPOSEDAD <- read.csv("Edad.csv", header = TRUE, encoding = "utf-8") str(GRUPOSEDAD)

GRUPOSEDAD <- subset(GRUPOSEDAD, Estado!= "TOTAL GLOBAL") GRUPOSEDAD <- subset(GRUPOSEDAD, sexo != "G" ) GRUPOSEDAD <- subset(GRUPOSEDAD, cortesedad != "Total")

str(GRUPOSEDAD)

entidades <- c('Aguascalientes', 'Baja California', 'Baja California Sur', 'Campeche', 'Chiapas', 'Chihuahua', 'Coahuila', 'Colima', 'Distrito Federal', 'Durango', 'Guanajuato', 'Guerrero', 'Hidalgo', 'Jalisco', 'MÃ©xico', 'MichoacÃ¡n', 'Morelos', 'Nayarit', 'Nuevo LeÃ³n', 'Oaxaca', 'Puebla', 'QuerÃ©taro', 'Quintana Roo', 'San Luis PotosÃ????', 'Sinaloa', 'Sonora', 'Tabasco', 'Tamaulipas', 'Tlaxcala', 'Veracruz', 'YucatÃ¡n', 'Zacatecas')

entidades

par(mfrow = c(4, 8), mar = c(0, 0, 1, 0), oma = c(0, 0, 1.2, 0))

anio <- 2013 GRUPOSEDAD.año <- subset(GRUPOSEDAD, año == anio) for(i in 1:32) { GRUPOSEDAD.entidad <- subset(GRUPOSEDAD.año, Estado == entidades[i])

with(GRUPOSEDAD.entidad, plot(0, 0, type = "n", xlim = c(-1000, 1000), ylim = c(0, 13), xlab ="", ylab = "", axes = FALSE, bty="n", cex.main = 0.5, main = entidades[i]) )

rect(0, 0:11, -(GRUPOSEDAD.entidad[1:12, ]$numerocasos), 1:12, col="blue", border="white") rect(0, 0:11, (GRUPOSEDAD.entidad[13:24, ]$numerocasos), 1:12, col="red", border="white") } # Incluimos el año en el márgen superior externo. mtext(paste0(anio, " (Hombres, Mujeres)"), side = 3, outer = TRUE)

# Conclusión