

# Examen de Admision Practico

Benjamin Lopez

6/4/2021

## Problema 1

Generar una tabla con 3 columnas: Fecha Confirmados de SARS-CoV2 en Sonora por fecha usando la fecha de inicio de síntomas (no acumulados) Decesos (entre los confirmados) por fecha.

Para resolver este problema se utiliza el archivo 200511COVID19MEXICO.csv indicado en el ejercicio, el cual contiene la información de los casos de covid en mexico hasta esa fecha, de igual manera se hace uso de los archivos Catalogos\_0412.xlsx y el archivo Descriptores\_0419.xlsx para dale una interpretacion a los datos.

En esta ocasión hare uso de R con RStudio para resolver el problema.

Pasos a seguir para la solucion del problema:

1.- importamos los datos en R, extrayendolos del archivo CSV “200511COVID19MEXICO.csv” ejecutando la siguiente línea de código;

```
datosNacionales<-read.csv("200511covid19mexico.csv",header=T ,sep=",")
```

datosNacionales es el vector en el que guardare la información extraída del CSV. utilizamos la funcion read.csv indicando el nombre del archivo. el parámetro header para indicar que la primera línea son los nombres de las columnas sep para indicar que el separador de los datos es la “,”.

2.- Posteriormente creamos un conjunto de datos con los nombres de las columnas que nos interesa extraer del vector:

```
header<- c("FECHA_SINTOMAS", "FECHA_DEF", "RESULTADO", "ENTIDAD_RES")
```

3.- ahora procedemos a extrear las columnas que necesitamos del vector datos nacionales

```
AuxiliarDatosNacionales<-subset(datosNacionales,select = header)
```

AuxiliarDatosNacionales almacenara la información extraída de datosNacionales con las columnas especificadas en el conjunto header.

Una vez hecho esto procedemos a extraer la información que nos pide el problema, pero primero tenemos que revisar el diccionario y catálogo de la base de datos

Se nos pide la fecha de inicio de sintomas que esta representada por la columna “FECHA\_SINTOMAS” la cual Identifica la fecha en que inició la sintomatología del paciente.

Para los casos confirmado hacemos uso de la columna RESULTADO la cual puede albergar 3 claves: 1,2,3 que significan Positivo,No Positivo y pendiente respectivamente.

Para identificar el estado de sonora encontramos que el codigo asignado para esta es el 26.

4.- por último para las fechas de defunción a menos que sea una fecha valida podemos interpretar que no ha habido deceso.

```
SonoraFallecidos<- subset(AuxiliarDatosNacionales,ENTIDAD_RES == 26 & RESULTADO == 1 & FECHA_DEF != "9999-99-99")
```

Eliminamos la columna entidad con la siguiente instrucción:

```
SonoraFallecidos2<-subset(SonoraFallecidos,select =c("FECHA_SINTOMAS","FECHA_DEF","RESULTADO"))
```

5.- Por ultimo exportamos la información al archivo Tabla1.CSV son la siguiente instrucción:

```
write.csv(SonoraFallecidos2,"Tabla1.csv")
```

## Problema 2

Generar una tabla con la cantidad de casos hospitalizados en los estados de Sonora, Chihuahua, Nuevo León y Puebla.

Para resolver este problema aprovechamos el mismo vector de datos del problema anterior: datosNacionales.

1.- Extraemos las columnas ENTIDAD\_RES y TIPO\_PACIENTE como se muestra a continuacion:

```
header2<-c("ENTIDAD_RES","TIPO_PACIENTE")
```

```
AuxiliarHospitalizados<-subset(datosNacionales, select =header2)
```

En este codigo estamos extrayendo las datos con las columnas ENTIDAD\_RES y TIPO\_PACIENTE y lo estamos guardando en un vector llamado AuxiliarHospitalizados.

posteriormente procedemos a revisar en el diccionario de datos y catalogo de datos las claves de los estados que se solicitan en el problema, siendo el resultado el siguiente:

8 para chihuahua 19 para nuevo leon 21 para puebla 26 para sonora

En el caso de la columna tipo de paciente obtenemos lo siguiente:

1 Ambulatorio 2 Hospitalizado 99 No Especificado.

Con este conocimiento procedemos a obtener las personas hospitalizadas en las entidades solicitadas de la siguiente manera:

```
hospitalizadosPorEntidades<-subset(AuxiliarHospitalizados, (ENTIDAD_RES == 8 | ENTIDAD_RES==19 | ENTIDAD_RES==21 | ENTIDAD_RES==26))
```

Y por ultimo procedemos a exportar los datos con el nombre tabla2.csv con la siguiente instruccion:

```
write.csv(hospitalizadosPorEntidades,"Tabla2.csv")
```

## Problema 3

Realizar dos gráficas y guardarlas en el repositorio como grafica1.png y grafica2.png, las cuales contengan lo siguiente: En grafica1 realizar una gráfica de barras verticales con el número de casos hospitalizados (de acuerdo a nuestro archivo) de los estados de Sonora, Chihuahua, Nuevo León y Puebla. En grafica2 graficar la serie de tiempo de confirmados a nivel Nacional en el tiempo.

Para realizar la primera grafica solicitada procedemos a utilizar el vector hospitalizadosPorEntidades, con la funcion Table agrupamos los datos por entidad y total de casos:

```
hospitalizadosPorEntidadesGrafica<- table(hospitalizadosPorEntidades$ENTIDAD_RES)
```

Creamos un conjunto para las etiquetas de la grafica con los nombres de los estados ya que solo tenemos referencias numericas de ellos:

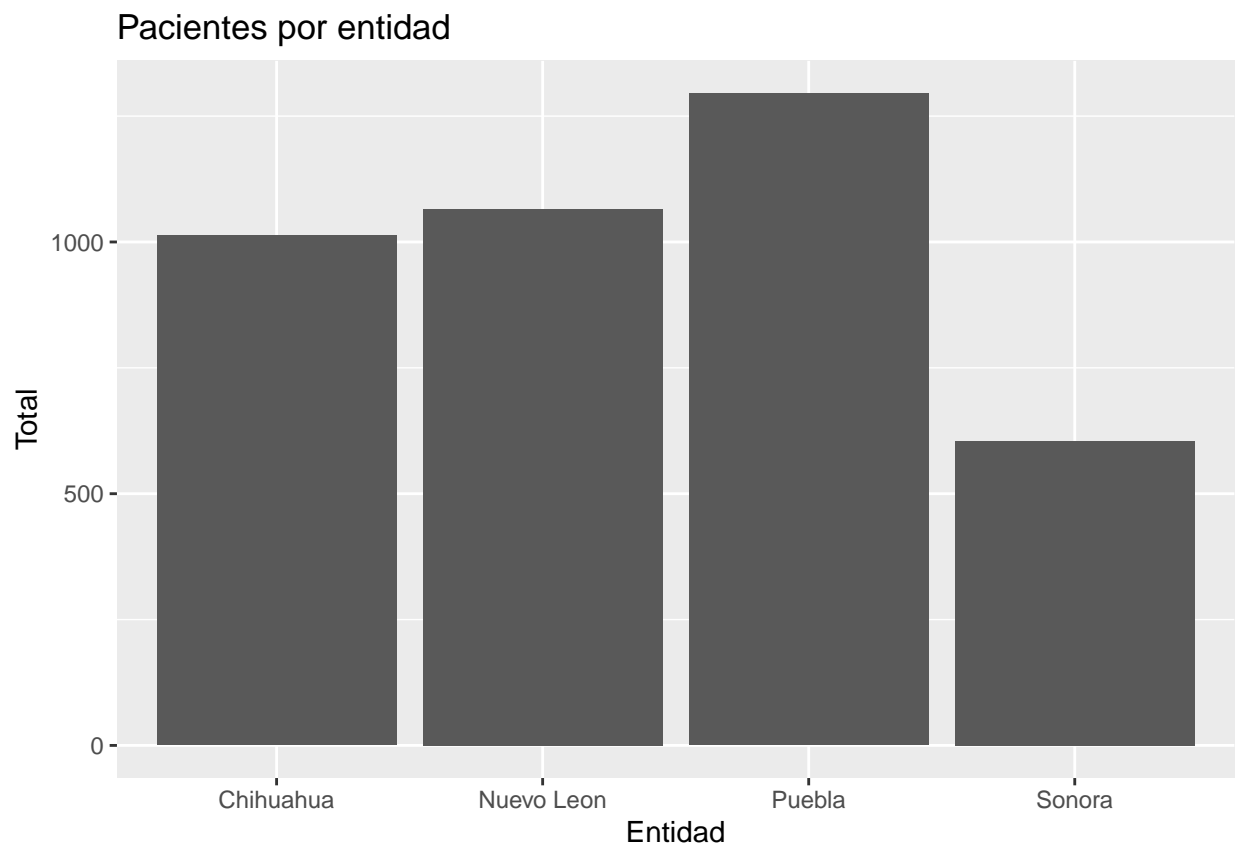
```
letras<-c("Chihuahua","Nuevo Leon","Puebla","Sonora")
```

Creamos un frame para graficar los datos en ggplot:

```
DatosAgrupadosEntidad<- data.frame(hospitalizadosPorEntidadesGrafica)
```

procedemos a realizar el grafico con la siguiente instruccion:

```
ggplot(DatosAgrupadosEntidad,aes(x=letras ,y=Freq))+ geom_col()+ labs(title = "Pacientes por entidad",x=letras)
```



Para la creacion de la grafica 2 extraemos 2 columnas de datosNacionales que son “FECHA\_SINTOMAS” y “RESULTADO” con la siguiente instruccion:

```
Nacional <- subset(datosNacionales, select = c("FECHA_SINTOMAS", "RESULTADO"))
```

Procedemos a extraer los datos positivos con la siguiente instruccion:

```
NacionalReducida <- subset(Nacional, RESULTADO == 1)
```

Convertimos en fecha la columna de FECHA\_SINTOMAS con la siguiente instruccion:

```
NacionalReducida$FECHA_SINTOMAS <- as.Date(NacionalReducida$FECHA_SINTOMAS)
```

Ahora procederemos a agrupar los datos por fecha y total de dias con la siguiente instruccion utilizando dplyr

```
TotalPorDia<-NacionalReducida%>%  
  select(FECHA_SINTOMAS,RESULTADO)%>%  
  group_by(FECHA_SINTOMAS)%>%  
  summarise(Total=sum(REULTADO))
```

con esto procedemos a graficar el resultado con la siguiente instruccion:

```
ggplot(TotalPorDia, aes(x=FECHA_SINTOMAS,y=Total))+ geom_line(colour="Blue")+labs(title = "Casos Sobre Tiempo")
```

