

Caso Estudio

Manejo de datos del covid

ISCIH

1 Preambulo

1.1 Objetivo

A partir de los datos de incidencia de COVID-19 en España (2020-2022) proporcionados por el área de vigilancia epidemiológica del CNE (<https://cnecovid.isciii.es/covid19>), se pretende describir la distribución de la epidemia entre grupos de edad y su evolución en el tiempo.

1.2 Importación de los datos

```
ruta.covid = "https://cnecovid.isciii.es/covid19/resources/casos_hosp_uci_def_sexo_edad_provres.csv"
# covid = read.table(ruta.covid, header=TRUE, sep=",") # con paquete base
require(data.table)
covid=fread(ruta.covid)
covid
```

```
##      provincia_iso sexo grupo_edad      fecha num_casos num_hosp num_uci
##      1:           A   H         0-9 2020-01-01         0         0         0
##      2:           A   H       10-19 2020-01-01         0         0         0
##      3:           A   H       20-29 2020-01-01         0         0         0
##      4:           A   H       30-39 2020-01-01         0         0         0
##      5:           A   H       40-49 2020-01-01         0         0         0
##      ---
## 1302206:          ZA  NC       50-59 2022-03-29         0         0         0
## 1302207:          ZA  NC       60-69 2022-03-29         0         0         0
## 1302208:          ZA  NC       70-79 2022-03-29         0         0         0
## 1302209:          ZA  NC        80+ 2022-03-29         0         0         0
## 1302210:          ZA  NC         NC 2022-03-29         0         0         0
##      num_def
##      1:      0
##      2:      0
##      3:      0
##      4:      0
##      5:      0
##      ---
## 1302206:      0
## 1302207:      0
## 1302208:      0
## 1302209:      0
## 1302210:      0
```

2 Formateo de la base

2.1 Creación de una variable temporal mensual

Redondear las fechas al mes (`?round.Date`).

```
covid[,periodo:=round(fecha,"months")]
```

2.2 Cambiar nombre de las variables

Eliminar el prefijo `num_` de las variables de conteo.

```
names(covid) <- gsub("^num_", "", names(covid))
```

2.3 Formato alargado de la base

Reestructurar la base de manera que la base tenga una fila por cada conteo :

```
covid.long = melt(covid,measure=c("casos","hosp","uci","def"))
covid.long
```

```
##      provincia_iso sexo grupo_edad      fecha      periodo variable value
##      1:           A   H         0-9 2020-01-01 2020-01-01     casos      0
##      2:           A   H        10-19 2020-01-01 2020-01-01     casos      0
##      3:           A   H        20-29 2020-01-01 2020-01-01     casos      0
##      4:           A   H        30-39 2020-01-01 2020-01-01     casos      0
##      5:           A   H        40-49 2020-01-01 2020-01-01     casos      0
##      ---
## 5208836:         ZA   NC        50-59 2022-03-29 2022-03-01       def      0
## 5208837:         ZA   NC        60-69 2022-03-29 2022-03-01       def      0
## 5208838:         ZA   NC        70-79 2022-03-29 2022-03-01       def      0
## 5208839:         ZA   NC          80+ 2022-03-29 2022-03-01       def      0
## 5208840:         ZA   NC          NC 2022-03-29 2022-03-01       def      0
```

3 Resumen de la base de datos

3.1 Resumen por grupos de edad y gravedad

A partir de la base anterior, crear una tabla de frecuencias con el grupo de edad en filas y la gravedad de la enfermedad en columnas:

```
dcast(covid.long,grupo_edad~variable,fun=sum)
```

```
##      grupo_edad  casos  hosp  uci  def
## 1:      0-9 1071081  5942  301   24
## 2:     10-19 1527859  4988  316   28
## 3:     20-29 1627666 15289  876  147
## 4:     30-39 1766228 29914 2328  354
```

```
## 5:      40-49 2091706  51732  5178  1173
## 6:      50-59 1494412  74921  9844  4005
## 7:      60-69  865327  89931 15058 10390
## 8:      70-79  548577  96415 12346 21955
## 9:       80+  529835 140208  2459 64037
## 10:      NC    9410   1580   161   206
```

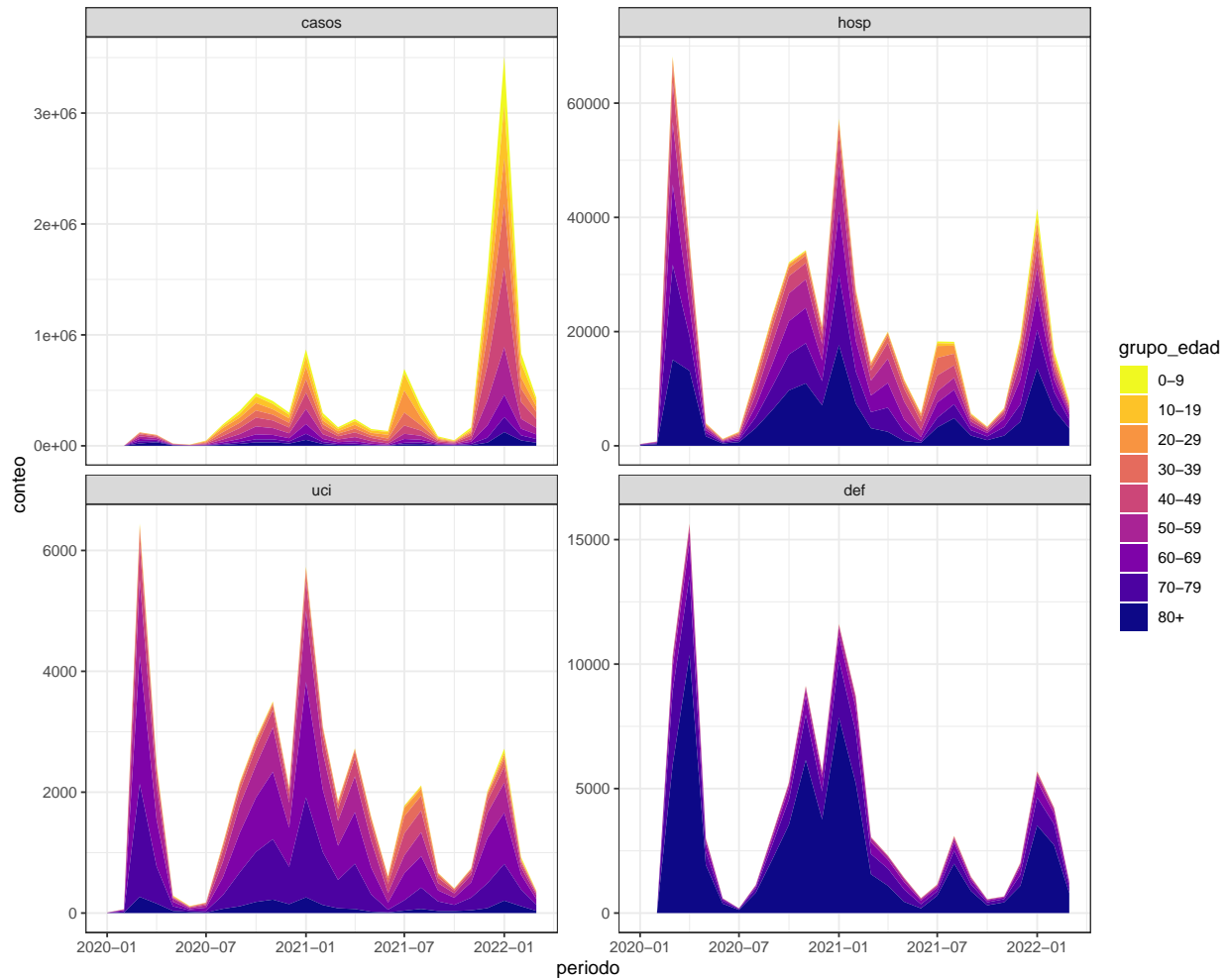
3.2 Resumen por grupos de edad, gravedad y periodo

Agregar la base de datos por grupo de edad (quitar registros sin información sobre la edad), periodo (mes) y gravedad.

```
resumen=covid.long[grupo_edad !="NC",.(conteo=sum(value)),by=.(grupo_edad,periodo,variable)]
resumen
```

```
##      grupo_edad      periodo variable conteo
## 1:      0-9 2020-01-01      casos      1
## 2:     10-19 2020-01-01      casos      0
## 3:     20-29 2020-01-01      casos      1
## 4:     30-39 2020-01-01      casos      1
## 5:     40-49 2020-01-01      casos      2
## ---
## 968:     40-49 2022-03-01      def      18
## 969:     50-59 2022-03-01      def      52
## 970:     60-69 2022-03-01      def     150
## 971:     70-79 2022-03-01      def     244
## 972:      80+ 2022-03-01      def     800
```

```
require(ggplot2)
ggplot(resumen,aes(y=conteo,x=periodo,fill=grupo_edad))+
  geom_area() +
  facet_wrap(variable~.,nrow=2,scales="free_y")+
  scale_fill_viridis_d(direction=-1,option="plasma")+
  scale_x_date()+
  theme_bw()
```



4 Juntar información demografica para cálculo de tasas

4.1 Descarga de los datos de población (fuente INE)

Descargar los datos de Población del INE:

```
ruta.ine="https://www.ine.es/jaxi/files/_px/es/csv_bdsc/t20/e245/p08/10/01002.csv_bdsc?nocab=1"
ine=fread(ruta.ine,encoding="UTF-8")
ine
```

```
##      Edad (grupos quinquenales) Españoles/Extranjeros      Sexo  Año
##      1:      TOTAL EDADES      TOTAL Ambos sexos 2021
##      2:      TOTAL EDADES      TOTAL Ambos sexos 2020
##      3:      TOTAL EDADES      TOTAL Ambos sexos 2019
##      4:      TOTAL EDADES      TOTAL Ambos sexos 2018
##      5:      TOTAL EDADES      TOTAL Ambos sexos 2017
##      ---
## 4748:      100 años y más      Extranjeros      Mujeres 2002
## 4749:      100 años y más      Extranjeros      Mujeres 2001
```

```
## 4750:          100 años y más          Extranjeros    Mujeres 2000
## 4751:          100 años y más          Extranjeros    Mujeres 1999
## 4752:          100 años y más          Extranjeros    Mujeres 1998
##          Total
##    1: 47.385.107
##    2: 47.450.795
##    3: 47.026.208
##    4: 46.722.980
##    5: 46.572.132
##    ---
## 4748:          199
## 4749:          276
## 4750:          239
## 4751:          189
## 4752:           46
```

4.2 Formateo de la base INE

Filtrar y formatear la base del INE para poder juntar los datos de población de 2021 (por grupos de edad) a la base agregada de covid anterior.

```
setnames(ine,1:5,c("grupo_edad_ine","origen","sexo","año","total"))
temp=subset(ine,año==2021 & sexo=="Ambos sexos" & grupo_edad_ine!="TOTAL EDADES" & origen=="TOTAL")
temp[,pob:=as.numeric(gsub("\\.", "", total))]

grupos.ine=c('0-4 años','5-9 años','10-14 años','15-19 años','20-24 años',
            '25-29 años','30-34 años','35-39 años','40-44 años','45-49 años',
            '50-54 años','55-59 años','60-64 años','65-69 años','70-74 años',
            '75-79 años','80-84 años','85-89 años','90-94 años','95-99 años',
            '100 años y más')

grupos.covid=c(rep(c("0-9","10-19","20-29","30-39",
                    "40-49","50-59","60-69","70-79"),each=2),rep("80+",5))

temp[,grupo_edad:=factor(grupo_edad_ine,levels=grupos.ine,labels=grupos.covid)]
temp
```

```
##      grupo_edad_ine origen      sexo año      total      pob grupo_edad
##  1:      0-4 años  TOTAL Ambos sexos 2021 1.887.689 1887689      0-9
##  2:      5-9 años  TOTAL Ambos sexos 2021 2.275.778 2275778      0-9
##  3:     10-14 años  TOTAL Ambos sexos 2021 2.526.140 2526140     10-19
##  4:     15-19 años  TOTAL Ambos sexos 2021 2.417.918 2417918     10-19
##  5:     20-24 años  TOTAL Ambos sexos 2021 2.382.448 2382448     20-29
##  6:     25-29 años  TOTAL Ambos sexos 2021 2.548.736 2548736     20-29
##  7:     30-34 años  TOTAL Ambos sexos 2021 2.786.314 2786314     30-39
##  8:     35-39 años  TOTAL Ambos sexos 2021 3.250.538 3250538     30-39
##  9:     40-44 años  TOTAL Ambos sexos 2021 3.923.707 3923707     40-49
## 10:     45-49 años  TOTAL Ambos sexos 2021 3.945.845 3945845     40-49
## 11:     50-54 años  TOTAL Ambos sexos 2021 3.699.275 3699275     50-59
## 12:     55-59 años  TOTAL Ambos sexos 2021 3.422.655 3422655     50-59
## 13:     60-64 años  TOTAL Ambos sexos 2021 3.007.236 3007236     60-69
## 14:     65-69 años  TOTAL Ambos sexos 2021 2.480.872 2480872     60-69
## 15:     70-74 años  TOTAL Ambos sexos 2021 2.199.993 2199993     70-79
```

```
## 16:      75-79 años  TOTAL Ambos sexos 2021 1.767.146 1767146      70-79
## 17:      80-84 años  TOTAL Ambos sexos 2021 1.294.253 1294253      80+
## 18:      85-89 años  TOTAL Ambos sexos 2021   991.099  991099      80+
## 19:      90-94 años  TOTAL Ambos sexos 2021   446.746  446746      80+
## 20:      95-99 años  TOTAL Ambos sexos 2021   112.699  112699      80+
## 21: 100 años y más  TOTAL Ambos sexos 2021    18.020   18020      80+
##      grupo_edad_ine origen      sexo año      total      pob grupo_edad
```

4.3 Juntar la base INE con la base agregada de covid

```
poblacion=temp[,.(pob=sum(pob)),by=grupo_edad]
todo=merge(resumen,poblacion,by="grupo_edad",all.x=TRUE)
todo
```

```
##      grupo_edad      periodo variable conteo      pob
## 1:      0-9 2020-01-01      casos      1 4163467
## 2:      0-9 2020-02-01      casos      1 4163467
## 3:      0-9 2020-03-01      casos     384 4163467
## 4:      0-9 2020-04-01      casos     287 4163467
## 5:      0-9 2020-05-01      casos     342 4163467
## ---
## 968:      80+ 2021-11-01      def      430 2862817
## 969:      80+ 2021-12-01      def     1094 2862817
## 970:      80+ 2022-01-01      def     3517 2862817
## 971:      80+ 2022-02-01      def     2729 2862817
## 972:      80+ 2022-03-01      def      800 2862817
```

4.4 Cálculo de las tasas

Calcular tasas de incidencia mensuales (por 100.000 habitantes) por grupos de edad y por categoría de gravedad.

$$\text{tasa} = \frac{\text{conteo}}{\text{poblacion}} \times 100.000$$

```
tasas=todo[,.(tasa=sum(conteo)/sum(pob)*100000),by=(grupo_edad,variable,periodo)]
```

4.5 Resumen de las tasas

Representar mediante una tabla el máximo anual de incidencia y mortalidad por COVID según grupos de edad:

```
tasas[,years:=year(periodo)]
maximo <- function(x) round(max(x),2)
tabla=dcast(tasas,grupo_edad~variable+years,value.var="tasa",
            fun=maximo,subset=(variable %in% c("casos","def")))
tabla
```

```
##      grupo_edad casos_2020 casos_2021 casos_2022 def_2020 def_2021 def_2022
## 1:      0-9      699.97    3264.85   10507.37     0.05     0.07     0.07
```

## 2:	10-19	1224.32	3482.83	9738.42	0.10	0.04	0.08
## 3:	20-29	1345.48	5175.49	8316.22	0.37	0.22	0.20
## 4:	30-39	1061.35	4416.03	9514.33	0.63	0.45	0.53
## 5:	40-49	1013.03	4017.66	9104.61	2.10	1.19	1.03
## 6:	50-59	1001.34	3141.66	6038.70	7.19	5.25	3.85
## 7:	60-69	805.72	2018.15	3665.82	24.94	17.75	11.70
## 8:	70-79	702.24	1387.75	3418.61	79.91	58.48	28.41
## 9:	80+	1110.48	1863.48	4311.28	362.23	274.24	122.85