

# Mentoría: Coronavirus en Argentina y el mundo

## Práctico 1: Análisis y Visualización

Matías A. Bettera Marcat

Lucas C. Didoné

Daniel Peralta

Patricia Gonzalez

29 de junio de 2020

### 1. Análisis estadístico de variables

#### 1.1. Usar distintos tipos de gráficos para describir sus distribuciones

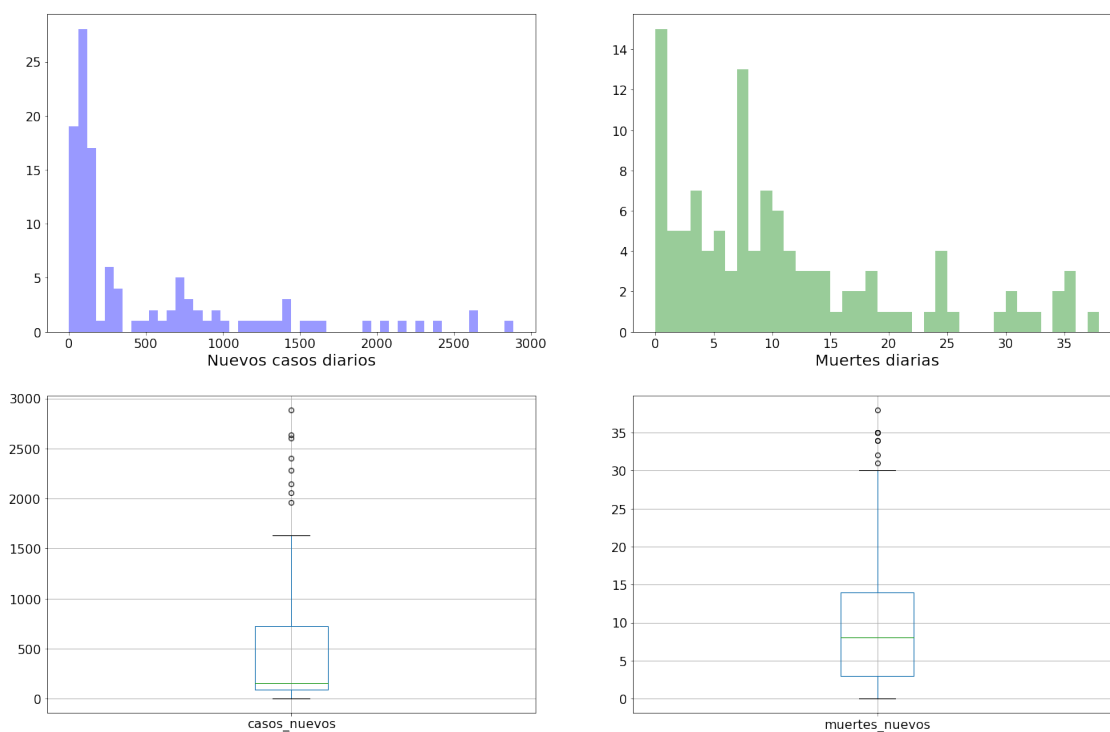


Figura 1: Arriba: histogramas de los valores diarios de nuevos diagnósticos de COVID-19 en Argentina (Izquierda) y de muertes por esta enfermedad en el país (Derecha). Abajo: diagramas de caja de la variables mostadas arriba.

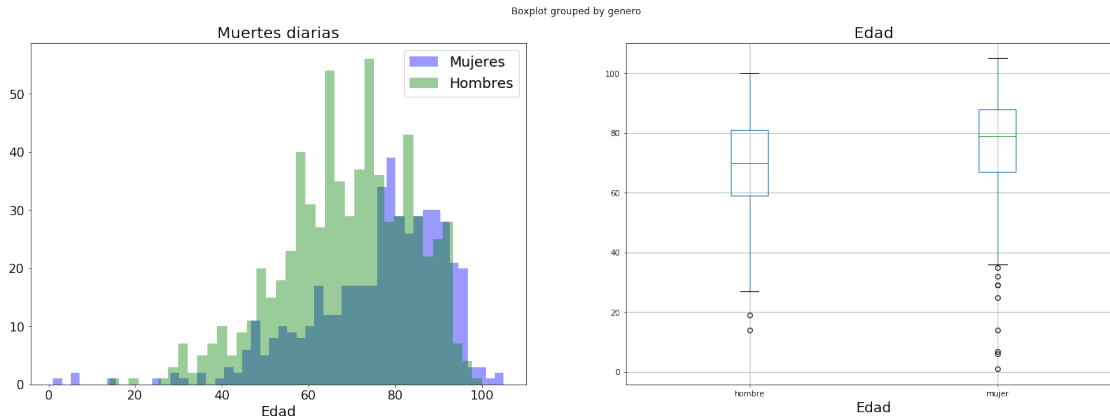


Figura 2: (Izquierda) Histograma de las edades de fallecidos por COVID-19 en Argentina. Se muestran separados por género. (Derecha) Diagramas de caja de las variables.

## 1.2. Analizar Outliers (si los hubiera)

En las figuras 1 se puede ver que la distribución de nuevos casos diarios tiene un pico para valores mucho menores que la media (moda = 88) y otro pico cercano a su media (media=502.2). Luego se ve que hay pocas ocurrencias aisladas de más de 2000 nuevos casos por día. Estos casos podrían ser considerados outliers, aunque la desviación estándar es grande ( $SD=665$ ). Un criterio usual para considerar outliers es que lo son casos que no estén contenidos en tres desviaciones estándar (hacia arriba o hacia abajo). Encontramos que hay sólo 3 son outliers con este criterio.

Por otro lado para las muertes diarias, no se observan gráficamente casos que se alejen mucho de la media y efectivamente con el criterio de tres desviaciones estándar no hay outliers.

En las distribuciones de edad de los fallecidos separadas por género, figura 2, se observa gráficamente que son bastante compactas y asimétricas con una cola más larga por la izquierda. Justamente por izquierda hay algunos casos aislados que se pueden considerar outliers, y con el criterio de tres desviaciones, vemos que son 5 casos de mujeres y 2 de hombres, de menores de 20 años.

En el jupyter notebook, se muestran las tablas de los outliers para esta cuatro variables analizadas.

## 1.3. Calcular estadísticos clásicos (media, mediana, moda, desviación estándar)

[32]:

	Variable	media	SD	mediana	moda	outliers
0	casos nuevos diarios	502.182609	664.904133	158.0	88.0	3
1	muertes diarias	10.469565	9.715359	8.0	0.0	0
2	edad de muerte de mujeres	75.776181	15.881272	79.0	88.0	5
3	edad de muerte de hombres	68.907010	15.054073	70.0	68.0	2

## 2. Evolución de variables en el tiempo



Figura 3: Evolución temporal de la variable '*casos\_nuevos*', que muestra los nuevos diagnósticos de COVID-19 diariamente.

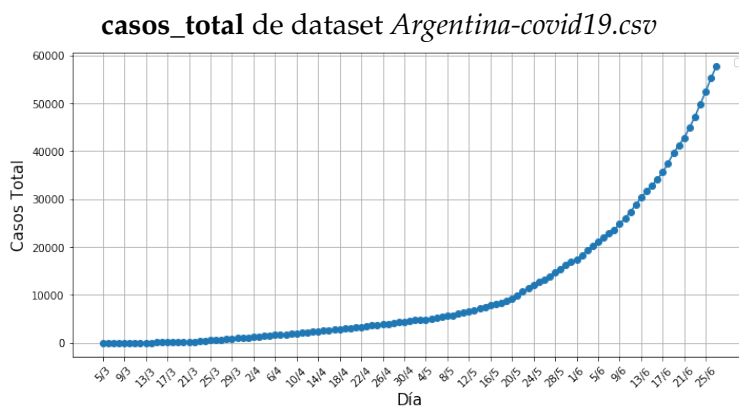


Figura 4: Evolución temporal del total de diagnósticos de COVID-19 en Argentina, variable '*casos\_total*'.

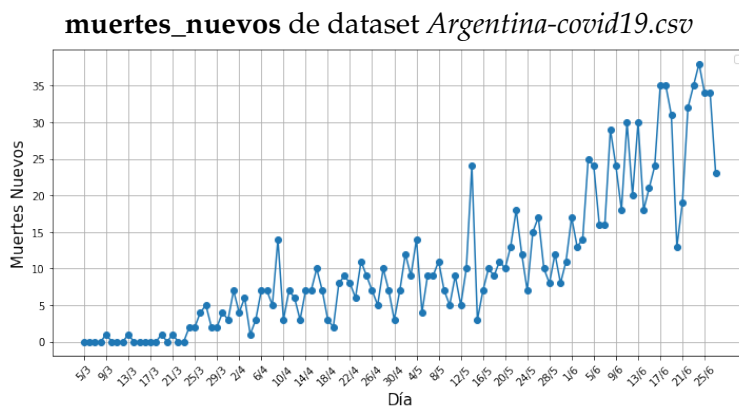


Figura 5: Evolución temporal de la variable '*muerteres\_nuevos*', que muestra los decesos del último día por COVID-19.

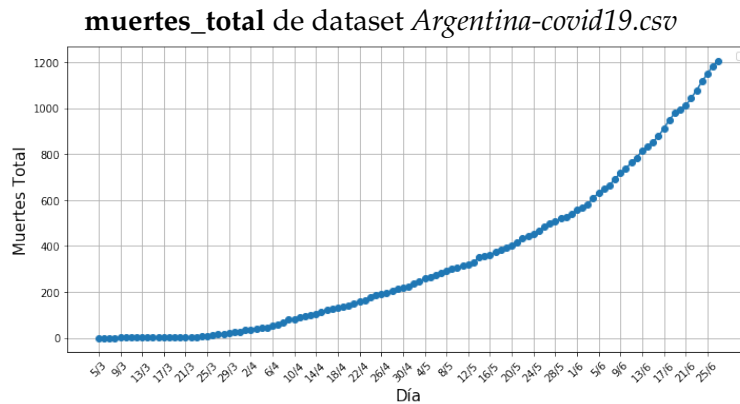


Figura 6: Evolución temporal del total decesos por COVID-19 en Argentina, variable 'muerres\_total'.

### 3. Análisis de probabilidades condicionales e independencia entre dos variables.

#### 3.1. ¿Cuál es la probabilidad de que un infectado sea mujer (varón)?

Probabilidad de que un infectado sea mujer: 0.494 .Porcentaje de mujeres infectadas: 49,4%

Probabilidad de que un infectado sea varón: 0.506 .Porcentaje de varones infectados: 50,6%

#### 3.2. ¿Cuál es la probabilidad de que un infectado sea de Córdoba (u otra provincia)?

Probabilidad de que un infectado sea de Buenos Aires: 0.481

Probabilidad de que un infectado sea de CABA: 0.425

Probabilidad de que un infectado sea de Chaco: 0.033

Probabilidad de que un infectado sea de Chubut: 0.002

Probabilidad de que un infectado sea de Corrientes: 0.002

Probabilidad de que un infectado sea de Córdoba: 0.011

Probabilidad de que un infectado sea de Entre Ríos: 0.004

Probabilidad de que un infectado sea de Formosa: 0.001

Probabilidad de que un infectado sea de Jujuy: 0.001

Probabilidad de que un infectado sea de La Pampa: 0.0

Probabilidad de que un infectado sea de La Rioja: 0.001

Probabilidad de que un infectado sea de Mendoza: 0.003

Probabilidad de que un infectado sea de Misiones: 0.001

Probabilidad de que un infectado sea de Neuquén: 0.007

Probabilidad de que un infectado sea de Río Negro: 0.014

Probabilidad de que un infectado sea de Salta: 0.0

Probabilidad de que un infectado sea de San Juan: 0.0

Probabilidad de que un infectado sea de San Luis: 0.0

Probabilidad de que un infectado sea de Santa Cruz: 0.001

Probabilidad de que un infectado sea de Santa Fe: 0.007

Probabilidad de que un infectado sea de Santiago del Estero: 0.0  
Probabilidad de que un infectado sea de Tierra del Fuego: 0.003  
Probabilidad de que un infectado sea de Tucumán: 0.001

### **3.3. Probabilidad de que un fallecido sea mujer (varón)**

Probabilidades de que un fallecido sea mujer u hombre:

$$P(\text{mujer}) = 0.41$$

$$P(\text{hombre}) = 0.59$$

### **3.4. Probabilidad de que un fallecido sea mujer (varón) dado que está en la franja etaria 60-80 (u alguna otra).**

\*Probabilidades de que un fallecido sea mujer u hombre, dado que se encontraban en

la franja etaria entre 60 y 80 años:

$$P(\text{mujer} \mid 60 < \text{edad} < 80) = 0.358$$

$$P(\text{hombre} \mid 60 < \text{edad} < 80) = 0.642$$

\*Probabilidades de que un fallecido sea mujer u hombre, y que se encontraban en la franja etaria entre 60 y 80 años:

$$P(60 < \text{edad} < 80) = 0.451$$

$$P(\text{mujer}, 60 < \text{edad} < 80) = 0.161$$

$$P(\text{hombre}, 60 < \text{edad} < 80) = 0.29$$

### **3.5. Probabilidad de que un fallecido en la franja etaria 60-80 sea mujer (varón).**

Probabilidad de que sea mujer de entre 60 y 80 años: 0.358

Probabilidad de que sea hombre de entre 60 y 80 años: 0.642

### **3.6. Probabilidad de que un fallecido sea de CABA dado que es mujer. Y viceversa, que sea mujer dado que es de CABA.**

Probabilidad de que sea de CABA dado que es mujer: 0.402

Probabilidad de que sea mujer dado que es de CABA: 0.431

### **3.7. Probabilidad de que un fallecido en la franja etaria 30-60 sea de Chaco. (Quizás encontremos alguna correlación interesante entre edad y provincia).**

Probabilidad de que sea de Chaco y tenga entre 30 y 60 años: 0.14

Probabilidad de que tenga entre 30 y 60 años dado que es de Chaco: 0.4

### 3.8. ¿El género de un fallecido es independiente de la franja etaria?

En este caso para verificar la independencia, comprobamos si cumple el siguiente criterio:  $P(A|B) = P(A)$ .

En nuestro caso  $P(A|B)$  es igual a `prob_fallec_mujer_60_80` (Probabilidad de que la fallecida sea mujer de entre 60 y 80 años) y  $P(A)$  es la probabilidad de que sea mujer.

```
[54]: prob_fallec_mujer_60_80 == prob_mujer
```

```
[54]: False
```

```
[55]: prob_fallec_hombre_60_80 == prob_varon
```

```
[55]: False
```

Comprobamos que no son independientes ya que no se cumple el criterio definido. Realizamos el mismo análisis para los varones y determinamos que tampoco se cumple el criterio. Podemos afirmar que el género de un fallecido no es independiente de la franja etaria.

### 3.9. ¿El género de un fallecido es independiente de la provincia?

En este caso para verificar la independencia, comprobamos si cumple el siguiente criterio:  $P(A \cap B) = P(A) * P(B)$ .

Dónde  $P(A \cap B)$  es `PROB_CHACO_MUJER` (Probabilidad de que sea del Chaco dado que es del Mujer), y  $P(A)$  es probabilidad que sea del Chaco y  $P(B)$  probabilidad que sea mujer.

```
[56]: PROB_CHACO = len(data3[(data3.provincia == 'Chaco')]) / len(data3)
```

```
[57]: PROB_CHACO_MUJER = len(data3[((data3.provincia == 'Chaco') & (data3.genero == 'mujer'))]) / len(data3[(data3.genero == 'mujer')])
```

```
[58]: PROB_CHACO_MUJER == prob_mujer * PROB_CHACO
```

```
[58]: False
```

Comprobamos que no son independientes ya que no se cumple el criterio definido.

### 3.10. ¿La franja etaria de un fallecido es independiente de la provincia?

En este caso para verificar la independencia, comprobamos si cumple el siguiente criterio:  $P(A \cap B) = P(A) * P(B)$ .

Dónde  $P(A \cap B)$  es `prob_CHACO_30_60` (Probabilidad de que sea de Chaco y tenga entre 30 y 60 años), y  $P(A)$  es probabilidad que sea del Chaco y  $P(B)$  probabilidad que tenga entre 30 y 60 años.

```
[59]: prob_30_60 = len(data3[((data3.edad >= 30) & (data3.edad <= 60))]) / len(data3)
```

```
[60]: prob_CHACO_30_60 == prob_30_60 * PROB_CHACO
```

[60]: False

Comprobamos que no son independientes ya que no se cumple el criterio definido.

### Fallecidos - Análisis de correlación entre edad y provincia.

A continuación se presenta un mapa de calor (figura 7) que permite visualizar los casos fallecidos agrupados por rangos de edad y ciudad. Se puede observar que para las provincias de Buenos Aires, Chaco y Córdoba, la mayor cantidad de fallecidos tienen entre 60 y 80 años; mientras que en CABA la mayor concentración de fallecidos se presenta en rangos etarios superiores 80 años.

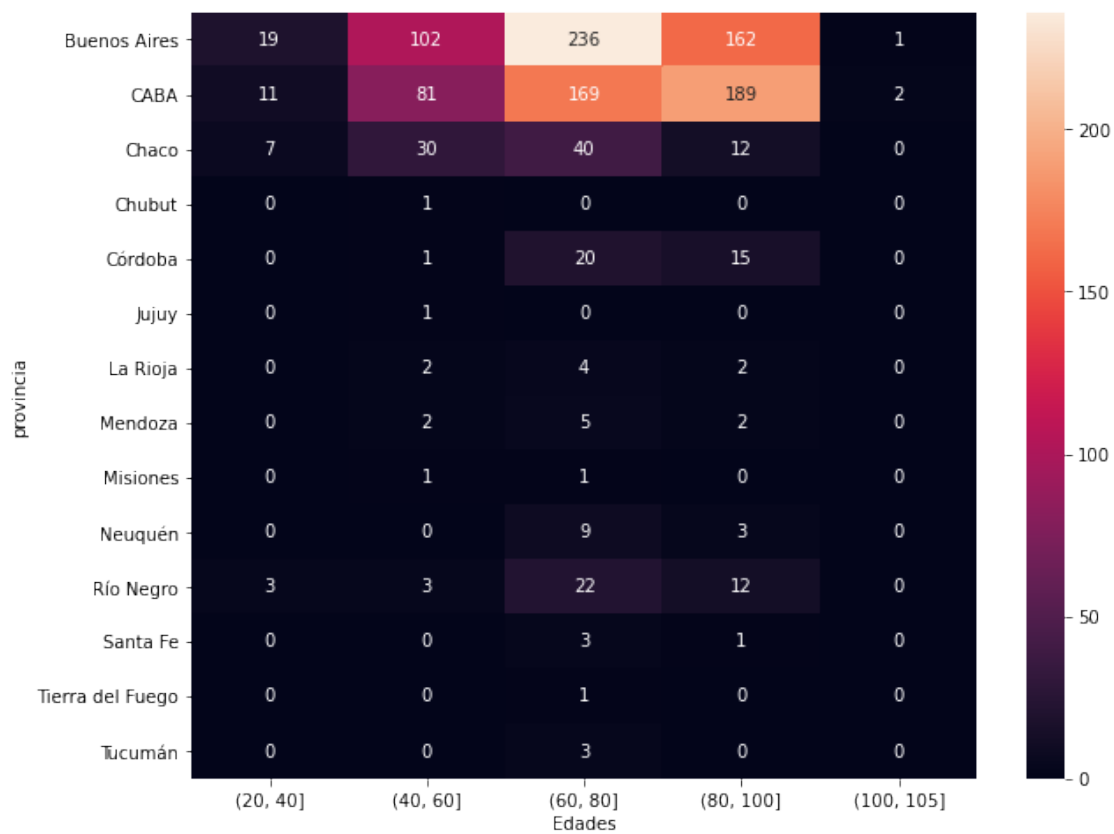


Figura 7: Mapa de correlación de fallecidos por provincia y grupo etáreo.