

COVID-19_SincronizadoWeb

March 22, 2020

Coronavirus COVID-19 .

(Año 2020)

Jesús Octavio García Flores

(j.octavio.ibs@gmail.com) En este pequeño ejercicio de análisis se plantean las siguientes preguntas:

```
# ¿Cuántos casos confirmados de personas infectadas hay en el mundo?
# ¿Qué país tiene el mayor número de casos?
# ¿Cuántas personas han fallecido por causa del coronavirus?
# ¿Qué porcentaje de la población infectada ha muerto?
# ¿Cómo ha sido el crecimiento de casos en un país determinado?
```

```
[1]: import pandas as pd
import matplotlib.pyplot as plt
%matplotlib inline
```

Full data set

```
[2]: source_full_data = "https://covid.ourworldindata.org/data/ecdc/full_data.csv"
data1 = pd.read_csv(source_full_data)
full_data = pd.DataFrame(data1)
full_data.head()
```

```
[2]:
```

	date	location	new_cases	new_deaths	total_cases	total_deaths
0	2019-12-31	Afghanistan	0	0	0	0
1	2020-01-01	Afghanistan	0	0	0	0
2	2020-01-02	Afghanistan	0	0	0	0
3	2020-01-03	Afghanistan	0	0	0	0
4	2020-01-04	Afghanistan	0	0	0	0

```
[3]: full_data[['new_cases', 'new_deaths', 'total_cases', 'total_deaths']].max()
```

```
[3]: new_cases      34047
      new_deaths     1690
      total_cases    305275
      total_deaths    12942
      dtype: int64
```

```
[4]: deaths = full_data['total_deaths'].max()
      print("\033[1m \n EL TOTAL DE MUERTES HASTA EL MOMENTO ES DE " + str(deaths) + "\n
      ↪" PERSONAS \033[0m \n")
```

EL TOTAL DE MUERTES HASTA EL MOMENTO ES DE 12942 PERSONAS

Total cases set

```
[5]: source_total_cases = "https://covid.ourworldindata.org/data/ecdc/total_cases.
      ↪csv"
      data2 = pd.read_csv(source_total_cases)
      total_cases = pd.DataFrame(data2)
      total_cases.head()
```

```
[5]:
```

	date	World	Afghanistan	Albania	Algeria	Andorra	Angola	\
0	2019-12-31	27	0.0	NaN	0.0	NaN	NaN	
1	2020-01-01	27	0.0	NaN	0.0	NaN	NaN	
2	2020-01-02	27	0.0	NaN	0.0	NaN	NaN	
3	2020-01-03	44	0.0	NaN	0.0	NaN	NaN	
4	2020-01-04	44	0.0	NaN	0.0	NaN	NaN	

	Antigua and Barbuda	Argentina	Armenia	...	United Arab Emirates	\
0	NaN	NaN	0.0	...	0.0	
1	NaN	NaN	0.0	...	0.0	
2	NaN	NaN	0.0	...	0.0	
3	NaN	NaN	0.0	...	0.0	
4	NaN	NaN	0.0	...	0.0	

	United Kingdom	United States	Uruguay	Uzbekistan	Vatican	Venezuela	\
0	0	0	NaN	NaN	NaN	NaN	
1	0	0	NaN	NaN	NaN	NaN	
2	0	0	NaN	NaN	NaN	NaN	
3	0	0	NaN	NaN	NaN	NaN	
4	0	0	NaN	NaN	NaN	NaN	

	Vietnam	Zambia	Zimbabwe
0	0.0	NaN	NaN
1	0.0	NaN	NaN

2	0.0	NaN	NaN
3	0.0	NaN	NaN
4	0.0	NaN	NaN

[5 rows x 181 columns]

```
[6]: #Número de casos
max_world = total_cases['World'].max()
```

```
[7]: print("\033[1m \n HASTA AHORA, HA HABIDO UN TOTAL DE " + str(max_world) + "\n
↪CASOS EN TODO EL MUNDO \033[0m \n")
print("\033[1m \n (El " + str(round((deaths/max_world)*100, 2)) + "% HA
↪FALLECIDO) \033[0m \n")
```

HASTA AHORA, HA HABIDO UN TOTAL DE 305275 CASOS EN TODO EL MUNDO

(El 4.24% HA FALLECIDO)

```
[8]: total_cases.describe()
```

```
[8]:
```

	World	Afghanistan	Albania	Algeria	Andorra	Angola	\
count	83.000000	73.000000	14.000000	78.000000	9.000000	1.0	
mean	59063.253012	2.452055	39.000000	8.717949	36.333333	2.0	
std	67908.741722	6.491535	25.534291	21.978087	35.958309	NaN	
min	27.000000	0.000000	2.000000	0.000000	1.000000	2.0	
25%	315.500000	0.000000	14.000000	0.000000	5.000000	2.0	
50%	40540.000000	0.000000	40.000000	0.000000	14.000000	2.0	
75%	88046.000000	0.000000	58.000000	1.000000	75.000000	2.0	
max	305275.000000	24.000000	76.000000	94.000000	88.000000	2.0	

	Antigua and Barbuda	Argentina	Armenia	Australia	...	\
count	3.0	16.000000	74.000000	83.000000	...	
mean	1.0	60.562500	9.972973	75.506024	...	
std	0.0	63.820549	32.086345	189.730899	...	
min	1.0	1.000000	0.000000	0.000000	...	
25%	1.0	11.250000	0.000000	0.000000	...	
50%	1.0	39.500000	0.000000	14.000000	...	
75%	1.0	83.500000	0.000000	27.500000	...	
max	1.0	225.000000	160.000000	1098.000000	...	

	United Arab Emirates	United Kingdom	United States	Uruguay	\
count	77.000000	83.000000	83.000000	8.000000	
mean	20.285714	296.385542	1175.650602	63.875000	

std	37.022346	871.621741	4094.993867	48.111292
min	0.000000	0.000000	0.000000	6.000000
25%	0.000000	0.000000	0.500000	23.750000
50%	5.000000	4.000000	12.000000	64.500000
75%	13.000000	29.500000	79.000000	98.000000
max	153.000000	5018.000000	26747.000000	135.000000

	Uzbekistan	Vatican	Venezuela	Vietnam	Zambia	Zimbabwe
count	7.000000	8.0	8.000000	79.000000	4.0	2.000000
mean	20.285714	1.0	28.625000	16.392405	2.0	1.500000
std	10.889051	0.0	10.126874	22.026451	0.0	0.707107
min	6.000000	1.0	10.000000	0.000000	2.0	1.000000
25%	12.000000	1.0	28.500000	0.000000	2.0	1.250000
50%	23.000000	1.0	33.000000	13.000000	2.0	1.500000
75%	28.000000	1.0	33.750000	16.000000	2.0	1.750000
max	33.000000	1.0	36.000000	94.000000	2.0	2.000000

[8 rows x 180 columns]

Total deaths

```
[9]: source_total_deaths = "https://covid.ourworldindata.org/data/ecdc/total_deaths.
    ↪CSV"
data3 = pd.read_csv(source_total_deaths)
total_deaths = pd.DataFrame(data3)
#total_deaths.head()
```

```
[10]: # Muertos en China a causa del coronavirus.
deaths_china = total_deaths['China'].max()
```

```
[11]: # Buscamos el país con mayor número de casos. Los datos son acumulados por día,
    ↪por eso buscamos el
    # máximo de cada país (porque es el acumulado total).
    # Indexamos con "loc" por el renglón "max" y hacemos la búsqueda a partir de la
    ↪columna "Afghanistan"

    # data_frame.idxmax(axis=1) devuelve el índice del elemento máximo por columna.

number_max_case = total_cases.describe().loc['max', 'Afghanistan'].max()
name_max_case = total_cases.describe().loc['max', 'Afghanistan'].idxmax(axis=1)

print("\033[1m \n EL PAÍS CON MAYOR NÚMERO DE CASOS HASTA EL MOMENTO ES ",
    ↪name_max_case, " CON ", number_max_case, " CASOS, \033[0m \n")

# Casos de muerte comparados con su población
porcentaje_deaths = (deaths_china/number_max_case)*100
```

```

print("\033[1m CON UN NÚMERO DE MUERTOS de", deaths_china , "PERSONAS (EL",
↳round(porcentaje_deaths,2), "% DE SU POBLACIÓN DE INFECTADOS). \033[0m \n")

# Casos totales comparados con los de China
porcentaje = (number_max_case / max_world)*100      # Porcentaje de infectados en
↳China
print("\033[1m ABARCANDO EL ", round(porcentaje,2), "% DEL TOTAL MUNDIAL.
↳\033[0m \n")

```

EL PAÍS CON MAYOR NÚMERO DE CASOS HASTA EL MOMENTO ES China CON 81499.0 CASOS,

CON UN NÚMERO DE MUERTOS de 3267 PERSONAS (EL 4.01 % DE SU POBLACIÓN DE INFECTADOS).

ABARCANDO EL 26.7 % DEL TOTAL MUNDIAL.

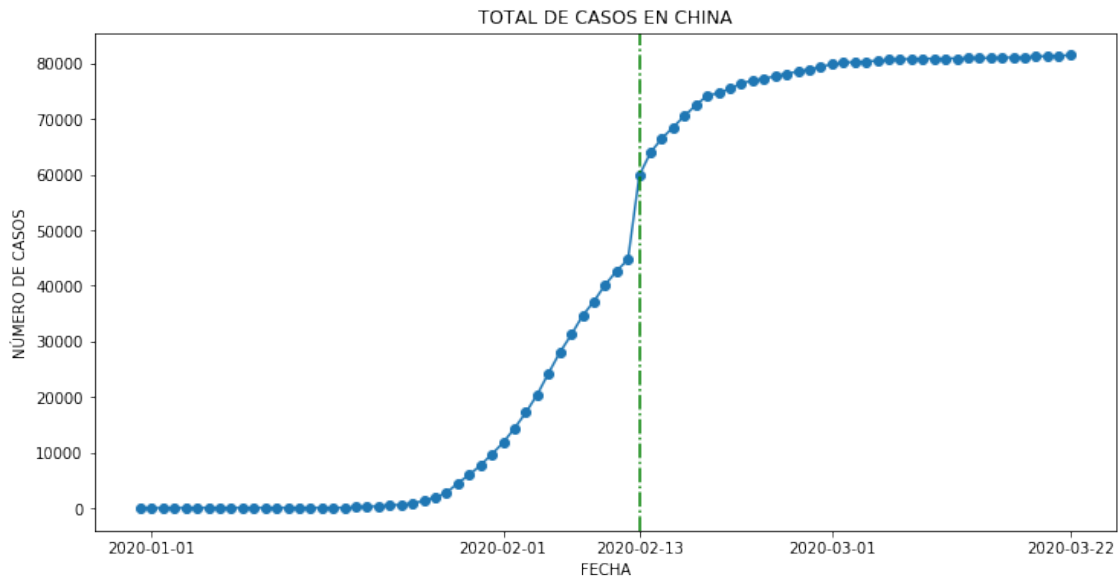
0.0.1 Enseguida se expone una gráfica del crecimiento que ha tenido la epidemia en este país.

```

[12]: plt.figure(figsize=(12,6))
plt.plot(total_cases['date'], total_cases['China'], marker='o')
plt.axvline(x='2020-02-13', color='g', ls='-.') #Grafica línea donde está el
↳máximo incremento

plt.xticks(['2020-01-01', '2020-02-01', '2020-02-13', '2020-03-01',
↳total_cases['date'][len(total_cases)-1]])
plt.xlabel('FECHA')
plt.ylabel('NÚMERO DE CASOS')
plt.title('TOTAL DE CASOS EN CHINA');

```



0.1 China ha disminuído las infecciones por contagio en los últimos días. Sus estrategias de contingencia le están dando muy buenos resultados.

Veamos un gráfico que representa la curva de los casos en China. Para esto, calculamos el número de infectados que aumentan por día en el país. Es decir, como los datos que tenemos son acumulados, hacemos la diferencia por cada día que pasa. Por ejemplo, si el primer día habían 10 casos y al segundo 15 casos, entonces hubo un aumento de 5 personas.

```
[13]: # Vamos a calcular los incrementos por día
china = total_cases[['date', 'China']]
china[0:5]
```

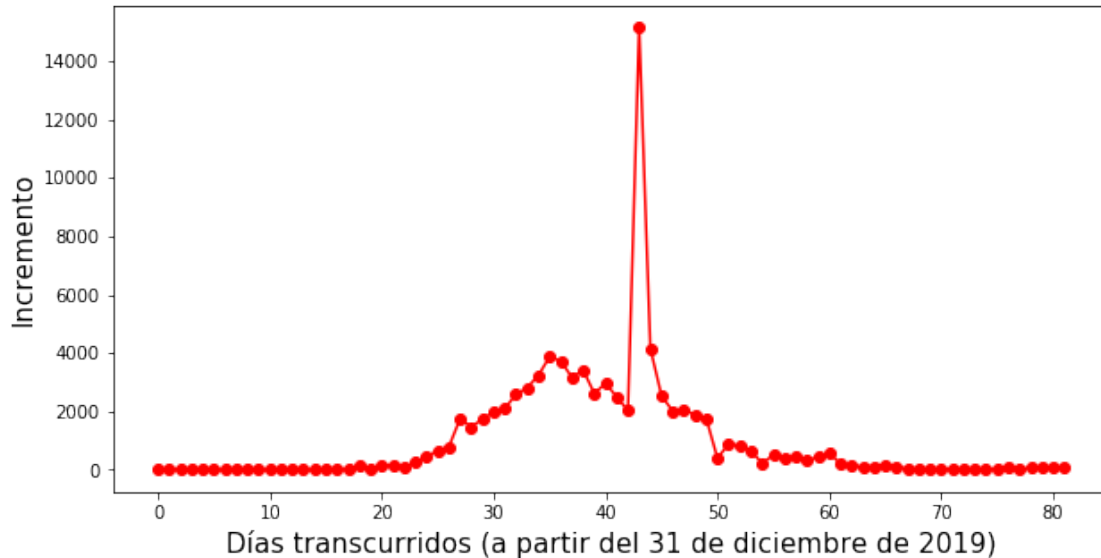
```
[13]:      date  China
0  2019-12-31    27
1  2020-01-01    27
2  2020-01-02    27
3  2020-01-03    44
4  2020-01-04    44
```

```
[47]: incremento = []

for n in range(1, len(china['China'])):
    diferencia = china['China'][n] - china['China'][n-1]
    incremento.append(diferencia)

#len(china['China']), len(incremento)
```

```
plt.figure(figsize=(10,5))
plt.plot(incremento, color='red', marker='o')
plt.xlabel("Días transcurridos (a partir del 31 de diciembre de 2019)", size=15)
plt.ylabel("Incremento", size=15);
```



```
[15]: print("\033[1m \n EL INCREMENTO MÁXIMO APARENTE FUE DE", max(incremento), "\n"
        ↪ "INFECCIONES. \033[0m \n")
```

EL INCREMENTO MÁXIMO APARENTE FUE DE 15141 INFECCIONES.

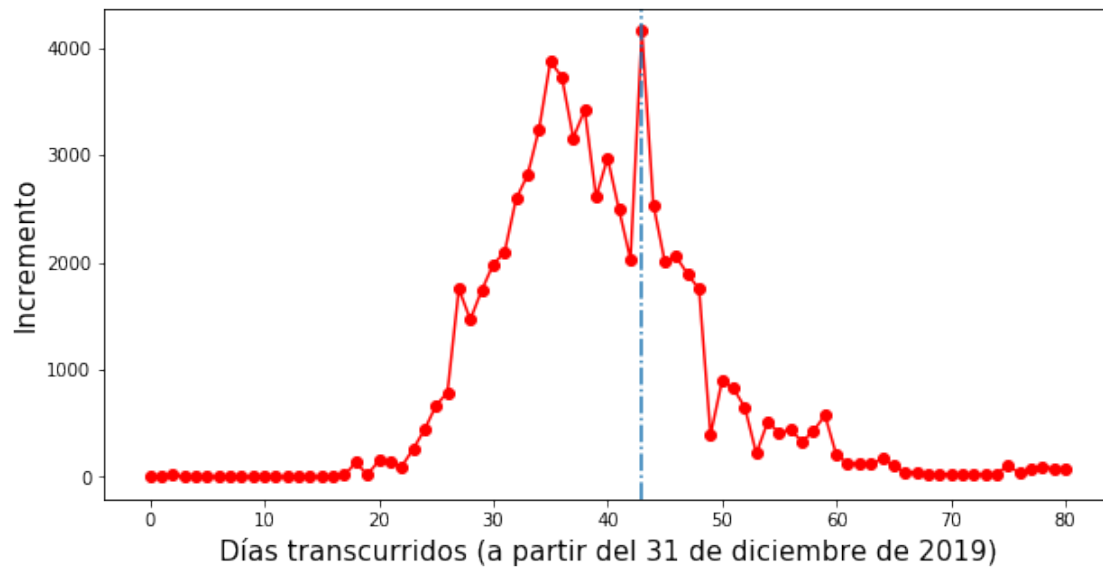
Este cambio intenso no es fidedigno, puesto que hubo un cambio en la forma de contar los casos en China por cuestiones de laboratorio. Si eliminamos los valores atípicos obtenemos lo siguiente:

```
[48]: def outlier_China():

        #Eliminamos el outlier (sin sustituirlo)
        incremento.pop(incremento.index(max(incremento)))

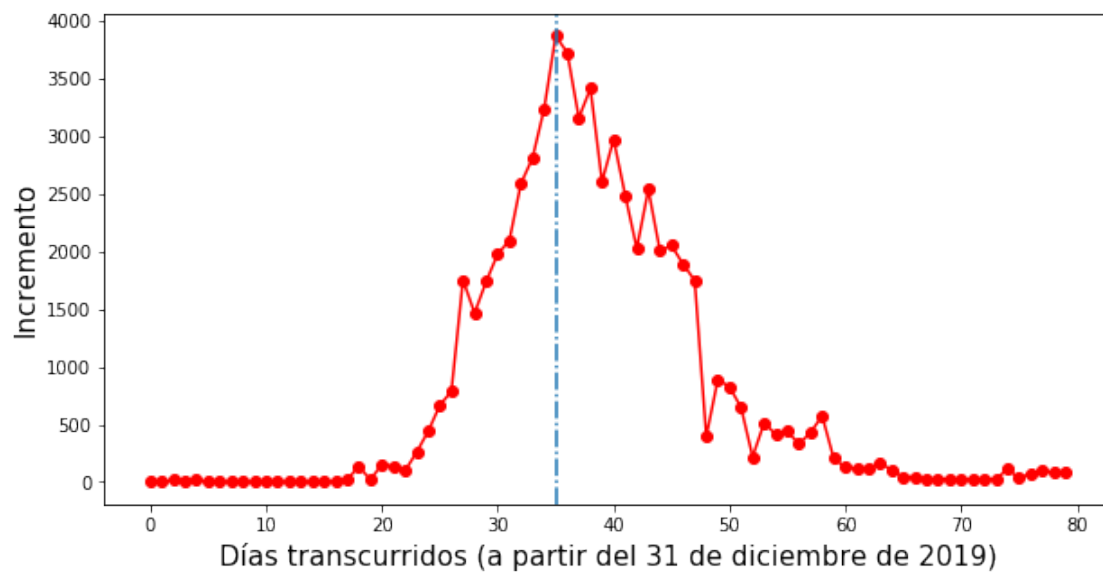
        #Calculamos nuevamente la coordenada x del nuevo valor máximo y graficamos.
        x_max = incremento.index(max(incremento))
        plt.figure(figsize=(10,5))
        plt.plot(incremento, color='red', marker='o')
        plt.axvline(x=x_max, ymin=0, ls='-.')
        plt.xlabel("Días transcurridos (a partir del 31 de diciembre de 2019)",
        ↪size=15)
        plt.ylabel("Incremento", size=15)
```

```
plt.show
outlier_China()
```



[49]: *# Parece que hay dos valores que se salen del rango abitual.
Así que volvemos a ajustar la gráfica quitando ese valor.*

```
outlier_China()
```




```
[50]: print("\033[1m \nEL INCREMENTO MÁXIMO FUE DE", max(incremento), "INFECCIONES_\n")
      ↪\033[0m \n")
```

EL INCREMENTO MÁXIMO FUE DE 3872 INFECCIONES

```
[51]: x_max = incremento.index(max(incremento)) #Devuelve el índice en la lista

print("\nEl maximo incremento se dio del ", china['date'][x_max-1], " al ",
      ↪china['date'][x_max], "\n")
```

El maximo incremento se dio del 2020-02-03 al 2020-02-04

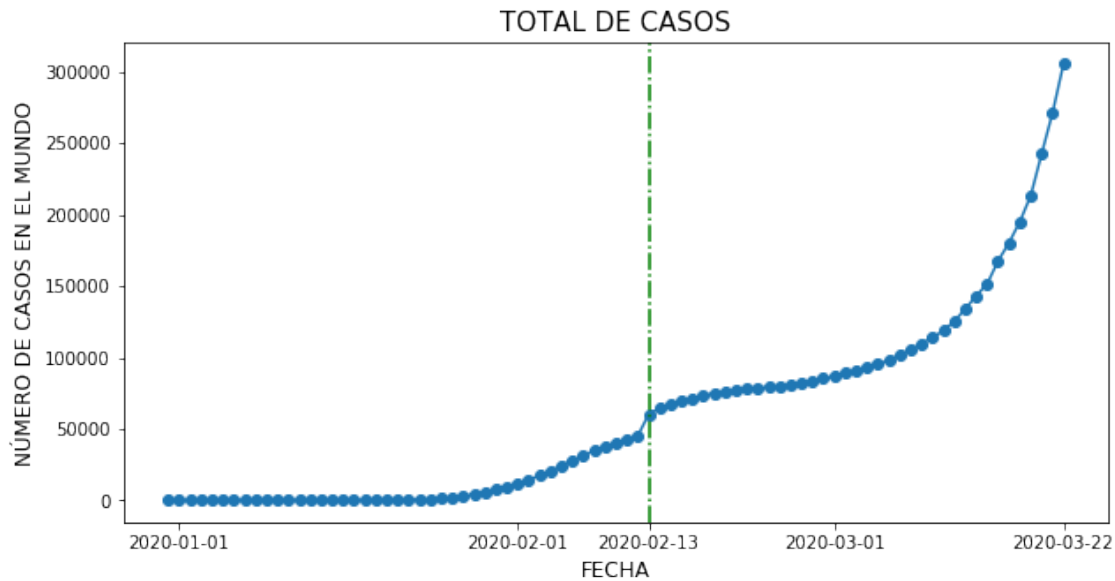
El primer incremento (17 nuevos infectados) se dio del 2 al 3 de enero 2020. El mayor incremento se dio del 3 al 4 de febrero.

0.2 Crecimiento mundial de la epidemia

```
[21]: cases_world = total_cases[['date', 'World']]
```

```
[53]: plt.figure(figsize=(10,5))
      plt.plot(cases_world['date'], cases_world['World'], marker='o')
      plt.axvline(x='2020-02-13', color='g', ls='-.')

      plt.xticks(['2020-01-01', '2020-02-01', '2020-02-13', '2020-03-01'],
      ↪cases_world['date'][len(total_cases)-1])
      plt.xlabel('FECHA', size=12)
      plt.ylabel('NÚMERO DE CASOS EN EL MUNDO', size=12)
      plt.title('TOTAL DE CASOS', size=15);
```



El crecimiento exponencial de infecciones asegura la categoría de PANDEMIA.

1 Por país

Hacemos una función que haga los cálculos respectivo por país, así como lo hicimos en el caso de

```
[54]: # Casos totales comparados con los del país elegido

def estadisticos(str_country):

    data_frame = total_cases.copy()

    max_world = data_frame['World'].max() # Total de casos mundial
    number_max_cases = data_frame.describe().loc['max',str_country] # Casos
    ↪ totales en el país elegido
    deaths_country = total_deaths[str_country].max()

    porcentaje = (number_max_cases / max_world)*100 # Porcentaje de infectados
    ↪ a nivel mundial

    porcentaje_deaths = (deaths_country/number_max_cases)*100 # Casos de
    ↪ muerte comparados con la población local

    #Grafica el número de casos diarios (acumulado)
    plt.figure(figsize=(10,5))
```

```

plt.plot(data_frame['date'], data_frame[str_country], marker='o')
plt.axvline(x='2020-02-13', color='g', ls='-.')

plt.xticks(['2020-01-01', '2020-02-01', '2020-02-13', '2020-03-01'],
→data_frame['date'][len(total_cases)-1])
plt.xlabel('Fecha', size=12)
plt.ylabel('Número de casos', size=12)
plt.title(f'Total de casos en {str_country}', size=15)
plt.show

# Cambia los NaN por ceros.
country = data_frame[['date', str_country]] #Subconjunto con los datos
→del país que necesitamos.
country = country.fillna(value = 0)

# Calcula los incrementos diarios
incremento = []
for n in range(1, len(country[str_country])):
    diferencia = country[str_country][n] - country[str_country][n-1]
    incremento.append(diferencia)

plt.figure(figsize=(10,5))

#Grafica una línea vertical donde se encuentra el máximo de la gráfica
x_max = incremento.index(max(incremento))
plt.axvline(x=x_max, ymin=0)

#Grafica la curva de incrementos
plt.plot(incremento, color='red', marker='o')
plt.xlabel('Días transcurridos (a partir del 31 de diciembre de 2019)',
→size=12)
plt.ylabel('Incremento', size=12)
plt.title(f'Incremento diario en {str_country}', size=15)
plt.show

print("\nEl número total de casos en", str_country, "es de",
→number_max_cases, "infectados.")

print("\nEl incremento máximo fue de ", max(incremento), "infecciones")

print("\nCon un número de muertos de", deaths_country, "personas (el",
→round(porcentaje_deaths,2), "% de su población infectada).\n")

print(str_country, "abarca el", round(porcentaje,2), "% del total de casos",
→confirmados a nivel mundial.")

```

2 Algunos países como ejemplo:

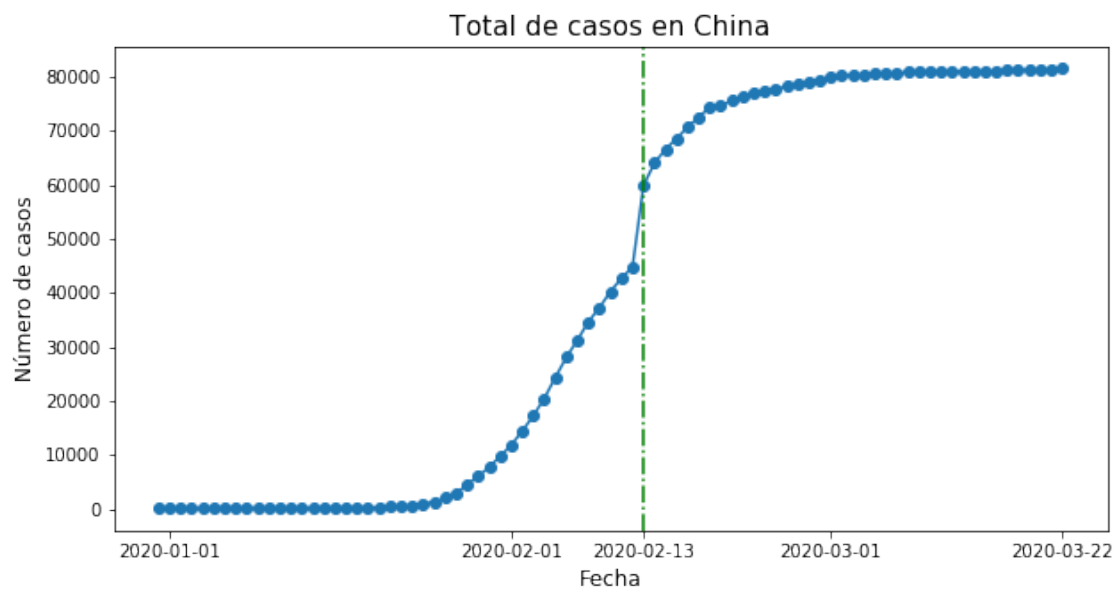
```
[55]: estadisticos('China')
```

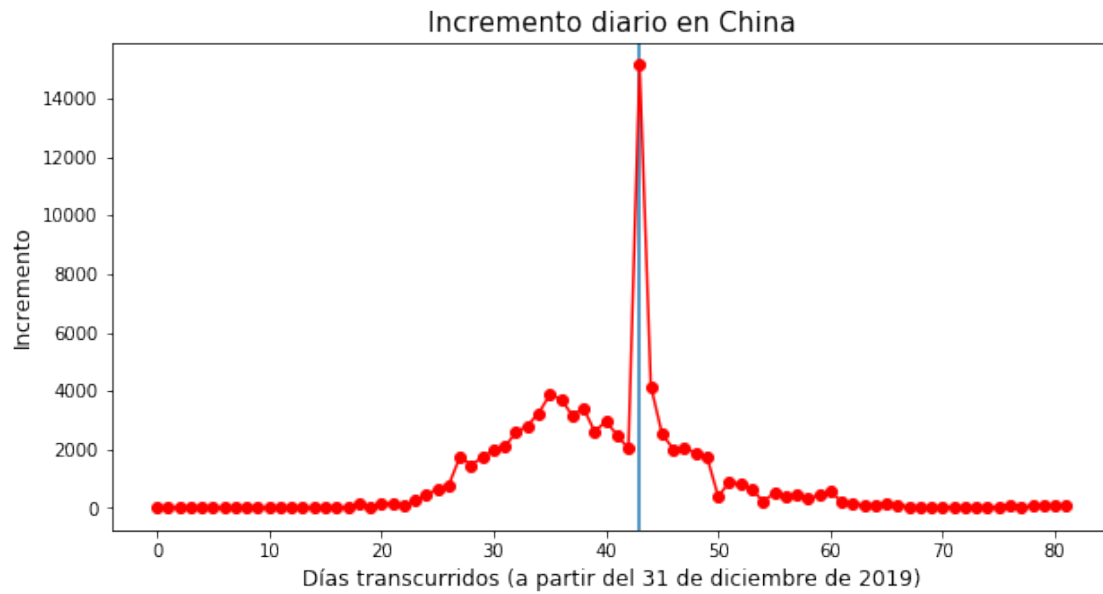
El número total de casos en China es de 81499.0 infectados.

El incremento máximo fue de 15141 infecciones

Con un número de muertos de 3267 personas (el 4.01 % de su población infectada).

China abarca el 26.7 % del total de casos confirmados a nivel mundial.





2.1 México

```
[56]: estadisticos('Mexico')
```

El número total de casos en Mexico es de 251.0 infectados.

El incremento máximo fue de 48.0 infecciones

Con un número de muertos de 2.0 personas (el 0.8 % de su población infectada).

Mexico abarca el 0.08 % del total de casos confirmados a nivel mundial.


```
0., 0., 0., 0., 0., 2., 4., 5., nan, nan, nan,
nan, nan, nan, 7., nan, nan, 11., 16., 26., 41., 53.,
82., 93., 118., 164., 203., 251.]])
```

2.1.1 Como puede verse, faltan registros para México. Pueden ajustarse con ayuda de la primer gráfica haciendo un seguimiento de su forma, pero apesar de ello, no cabe duda que el crecimiento es exponencial. Y podríamos decir (a groso modo) que se esperarían entre 50 y 200 casos nuevos de infectados del 20 al 21 de marzo, ya que ha tenido un control escaso (y notemos que la información que ofrece la OMS ha tenido por lo menos un día de retraso).

2.1.2 En adición a lo anterior, hasta la fecha de hoy (22 de marzo, 07:00 hrs), la estimación sobre los 50 nuevos casos mínimo esperados es muy buena, pues se registraron 48.

2.2 Canadá

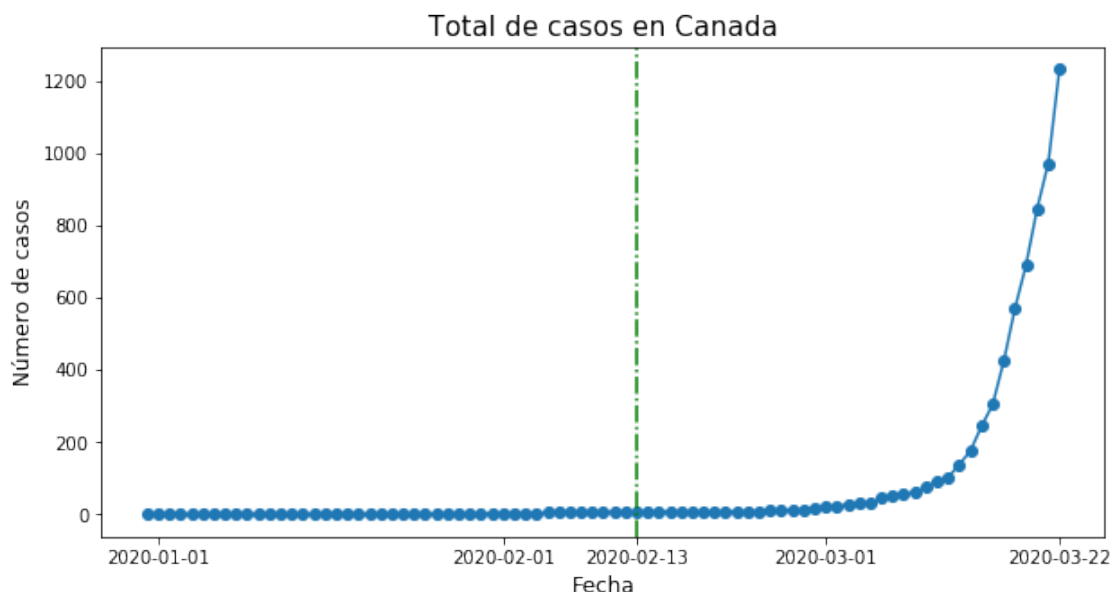
```
[58]: estadisticos('Canada')
```

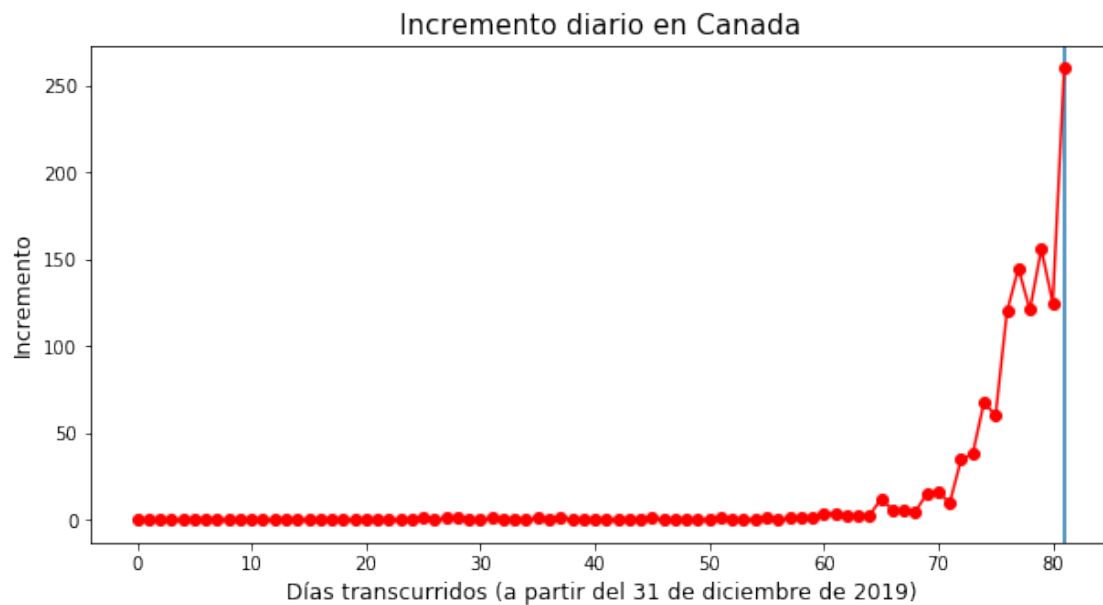
El número total de casos en Canada es de 1231.0 infectados.

El incremento máximo fue de 260 infecciones

Con un número de muertos de 13 personas (el 1.06 % de su población infectada).

Canada abarca el 0.4 % del total de casos confirmados a nivel mundial.





2.3 Estados Unidos

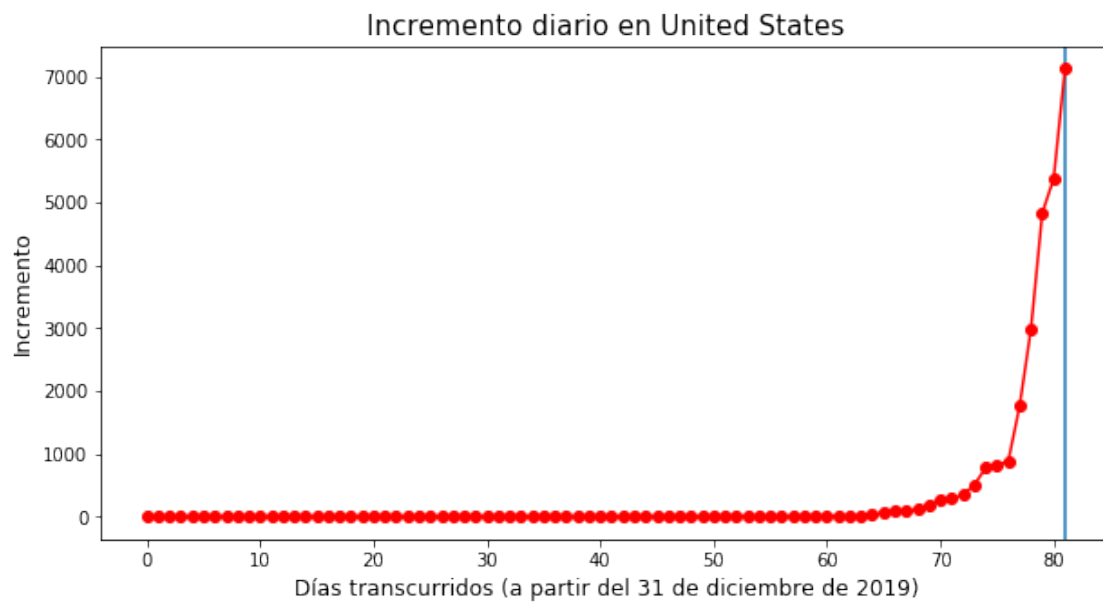
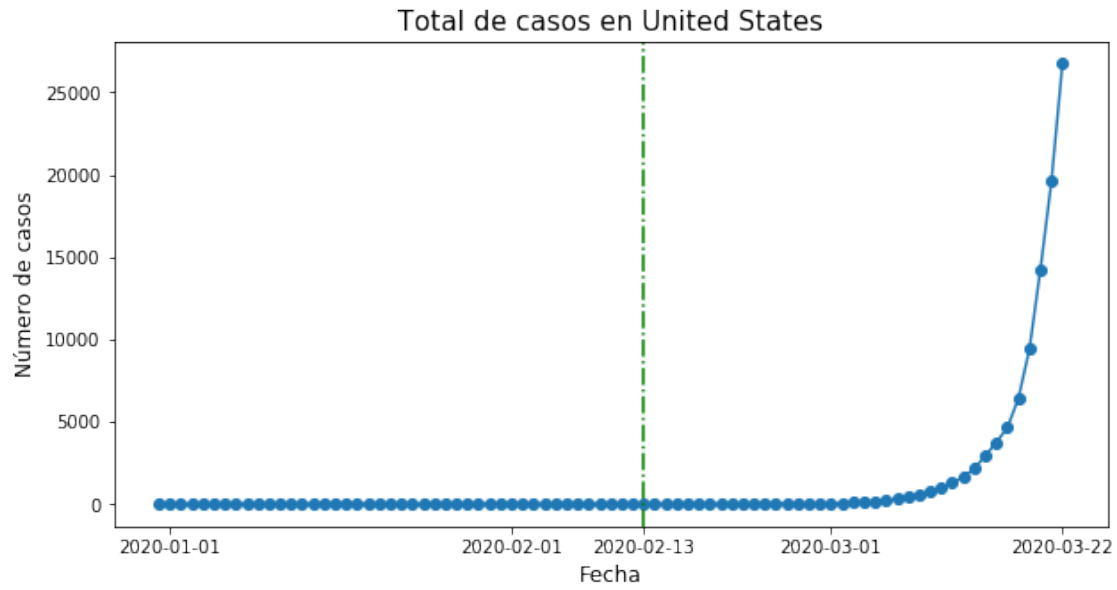
```
[59]: estadisticos('United States')
```

El número total de casos en United States es de 26747.0 infectados.

El incremento máximo fue de 7123 infecciones

Con un número de muertos de 340 personas (el 1.27 % de su población infectada).

United States abarca el 8.76 % del total de casos confirmados a nivel mundial.



2.4 Japón

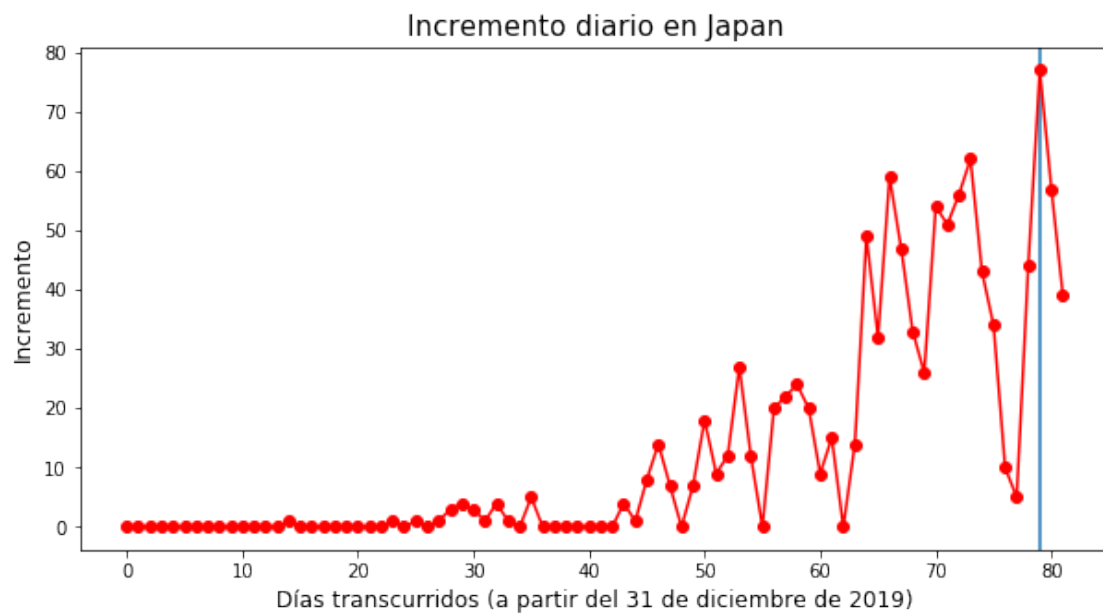
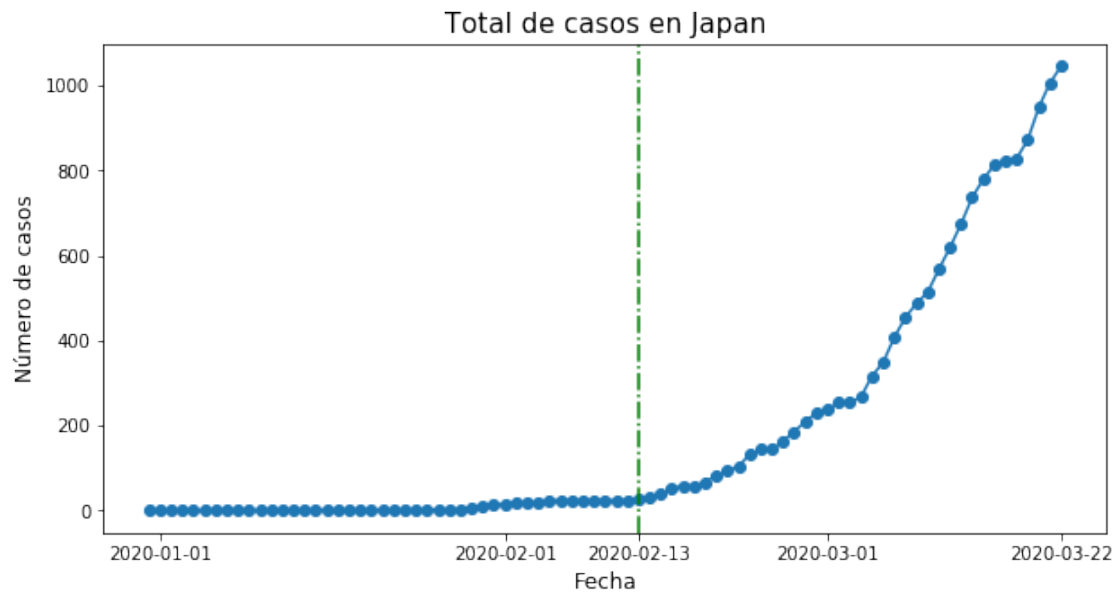
```
[60]: estadisticos('Japan')
```

El número total de casos en Japan es de 1046.0 infectados.

El incremento máximo fue de 77 infecciones

Con un número de muertos de 36 personas (el 3.44 % de su población infectada).

Japan abarca el 0.34 % del total de casos confirmados a nivel mundial.



2.5 España

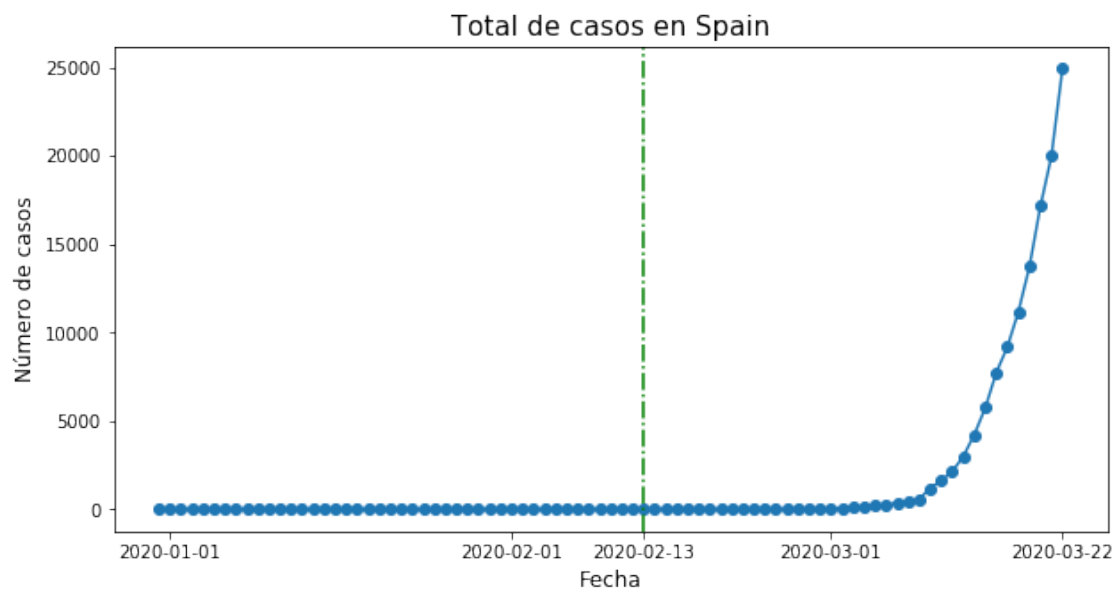
```
[61]: estadisticos('Spain')
```

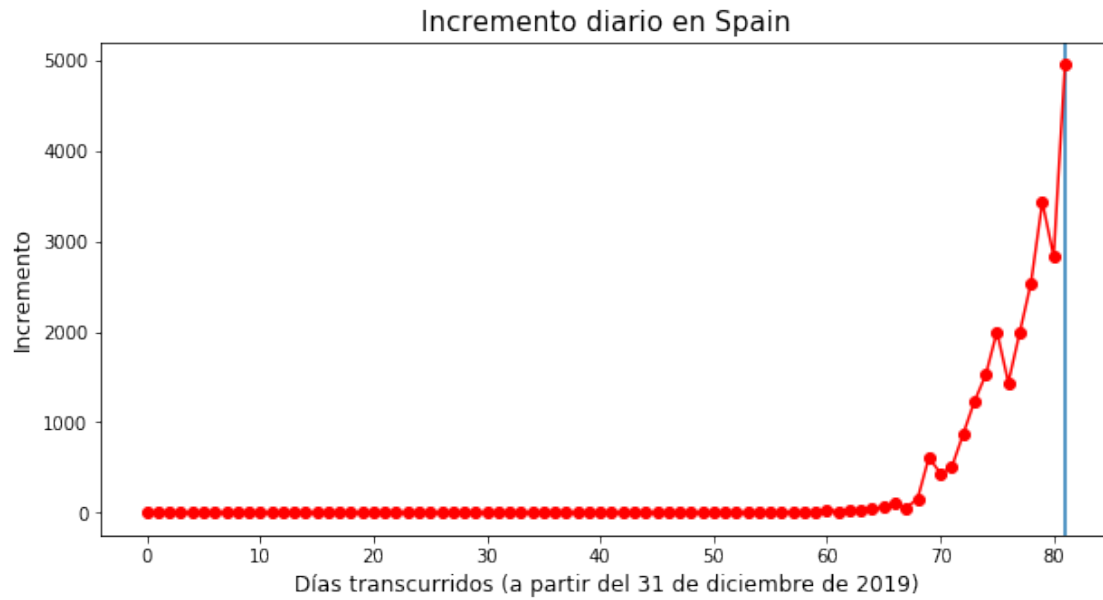
El número total de casos en Spain es de 24926.0 infectados.

El incremento máximo fue de 4946 infecciones

Con un número de muertos de 1326 personas (el 5.32 % de su población infectada).

Spain abarca el 8.17 % del total de casos confirmados a nivel mundial.





2.6 Italia

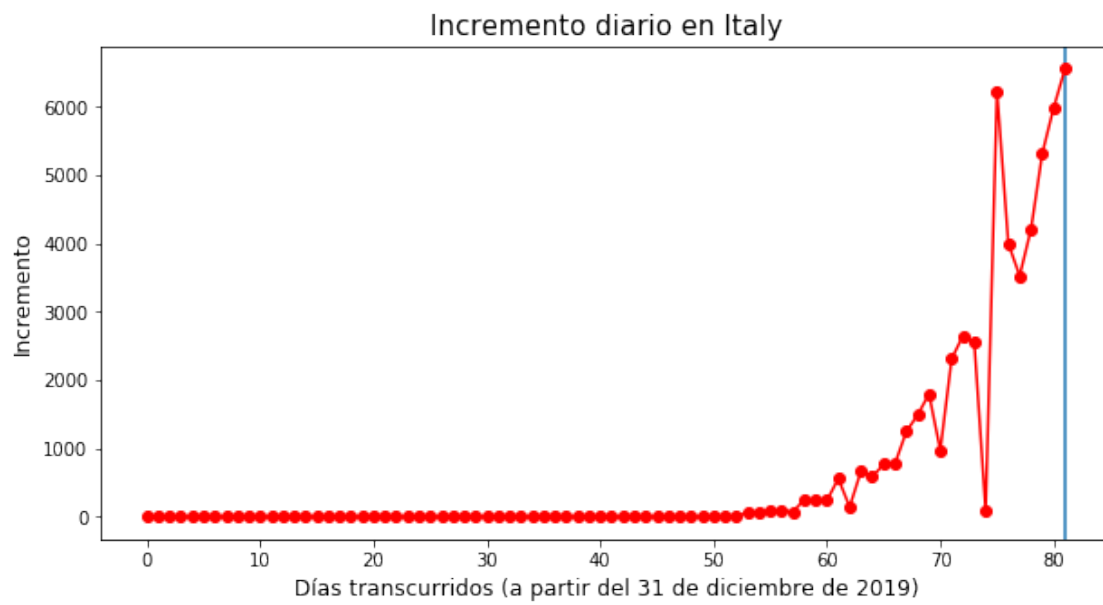
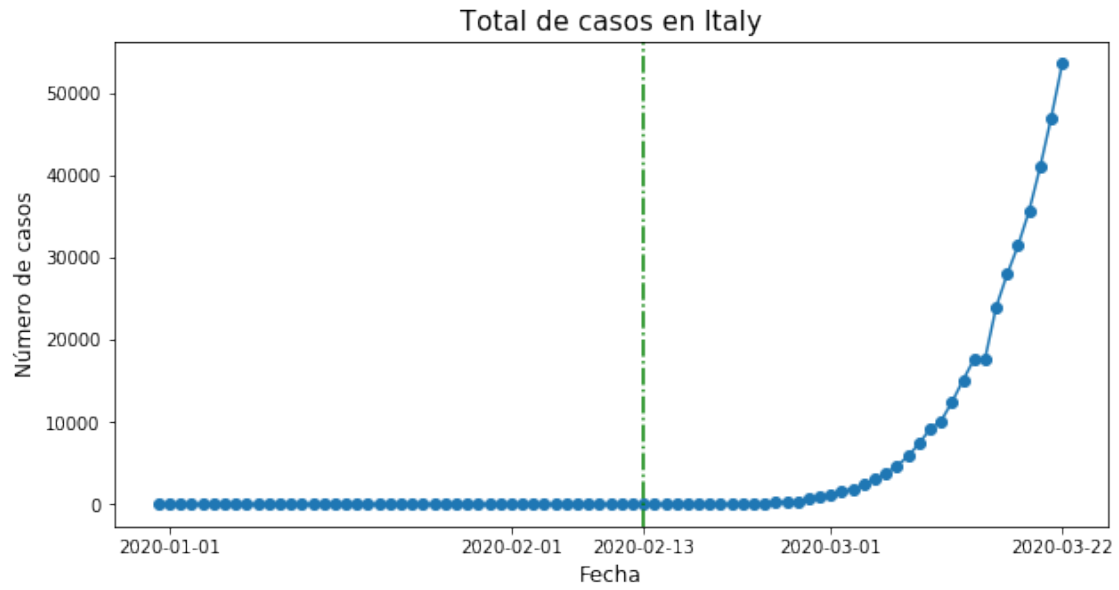
```
[62]: estadisticos('Italy')
```

El número total de casos en Italy es de 53578.0 infectados.

El incremento máximo fue de 6557 infecciones

Con un número de muertos de 4827 personas (el 9.01 % de su población infectada).

Italy abarca el 17.55 % del total de casos confirmados a nivel mundial.



[]: