

Model wirusa SARS-COV2

September 10, 2020

Autorzy: Jakub Jurczak, Szymon Kuliński, Piotr Głownia, Janusz Utrata, Bartek Sroczyński

1 Wstęp

W tym notebook-u prezentujemy nasz sposób przeprowadzenia analizy pandemii dla województwa małopolskiego. Ograniczymy się do użycia funkcji wykładniczej, w celu zrealizowania możliwego modelu predykcji. Autorzy zdają sobie sprawę, że istnieje ogromna liczba modeli, możliwych do użycia w regresji danych. Aczkolwiek naszym celem jest ukazanie najbardziej idealnego przypadku. Pozwoli nam to ukazać dynamikę modelu.

2 Analiza i implementacja

Import the necessary packages for data analysis.

```
[31]: import numpy as np
import pandas as pd
import matplotlib.pyplot as plt
import scipy
import sklearn
from scipy.optimize import curve_fit
from sklearn.metrics import r2_score
```

Load the data and have a first glance at it.

```
[32]: try:
    csv_path = "../COVID19-modelling/data/covid_pol_voivodeship.csv"
    load_csv = pd.read_csv(csv_path)
    print("The data was loaded correctly")
except NameError:
    print("The path", csv_path, "cannot be found")

load_csv
```

The data was loaded correctly

```
[32]:
```

	Date (CET) \
0	4 March 2020
1	5 March 2020

2	6 March 2020
3	7 March 2020
4	8 March 2020
..	...
176	27 August 2020
177	28 August 2020
178	Infections per voivodeship
179	Date (CET)
180	Notes: ^ Laboratory-confirmed cases, includi...

	Lower Silesia (DS) \
0	NaN
1	NaN
2	1
3	NaN
4	1
..	...
176	28
177	30
178	3995
179	Lower Silesia (DS)
180	Notes: ^ Laboratory-confirmed cases, includi...

	Kuyavia-Pomerania (KP) \
0	NaN
1	NaN
2	NaN
3	NaN
4	NaN
..	...
176	24
177	38
178	1268
179	Kuyavia-Pomerania (KP)
180	Notes: ^ Laboratory-confirmed cases, includi...

	Lubusz (LB) \
0	1
1	NaN
2	NaN
3	NaN
4	NaN
..	...
176	12
177	9
178	660
179	Lubusz (LB)

180 Notes: ^ Laboratory-confirmed cases, includi...

	Łódź (LD) \
0	NaN
1	NaN
2	NaN
3	NaN
4	NaN
..	...
176	47
177	39
178	5196
179	Łódź (LD)

180 Notes: ^ Laboratory-confirmed cases, includi...

	Lublin (LU) \
0	NaN
1	NaN
2	NaN
3	NaN
4	NaN
..	...
176	14
177	27
178	1484
179	Lublin (LU)

180 Notes: ^ Laboratory-confirmed cases, includi...

	Lesser Poland (MA) \
0	NaN
1	NaN
2	NaN
3	NaN
4	NaN
..	...
176	233
177	139
178	7128
179	Lesser Poland (MA)

180 Notes: ^ Laboratory-confirmed cases, includi...

	Masovia (MZ; Warsaw) \
0	NaN
1	NaN
2	NaN
3	NaN
4	2

```

..
176
177
178
179
180 Notes: ^ Laboratory-confirmed cases, includi...

                                Opole (OP) \
0                                NaN
1                                NaN
2                                NaN
3                                NaN
4                                NaN
..
176
177
178
179
180 Notes: ^ Laboratory-confirmed cases, includi...

                                Opole (OP) \
0                                NaN
1                                NaN
2                                NaN
3                                NaN
4                                NaN
..
176
177
178
179
180 Notes: ^ Laboratory-confirmed cases, includi...

                                Podlaskie (PD) \
0                                NaN
1                                NaN
2                                NaN
3                                NaN
4                                NaN
..
176
177
178
179
180 Notes: ^ Laboratory-confirmed cases, includi...

                                Podlaskie (PD) \
0                                NaN
1                                NaN
2                                NaN
3                                NaN
4                                NaN
..
176
177
178
179
180 Notes: ^ Laboratory-confirmed cases, includi...

                                Subcarpathian (PK) \
0                                NaN
1                                NaN
2                                NaN
3                                NaN
4                                NaN
..
176
177
178
179
180 Notes: ^ Laboratory-confirmed cases, includi...

                                Subcarpathian (PK) \
0                                NaN
1                                NaN
2                                NaN
3                                NaN
4                                NaN
..
176
177
178
179
180 Notes: ^ Laboratory-confirmed cases, includi...

                                Pomerania (PM) \
0                                NaN
1                                NaN
2                                NaN
3                                NaN
4                                NaN
..
176
177
178
179
180 Notes: ^ Laboratory-confirmed cases, includi...

```

0	NaN
1	NaN
2	NaN
3	NaN
4	NaN
..	...
176	71
177	42
178	2371
179	Pomerania (PM)
180	Notes: ^ Laboratory-confirmed cases, includi...

Holy Cross (SK) \

0	NaN
1	NaN
2	NaN
3	NaN
4	NaN
..	...
176	18
177	19
178	1380
179	Holy Cross (SK)
180	Notes: ^ Laboratory-confirmed cases, includi...

Silesia (SL) \

0	NaN
1	NaN
2	NaN
3	NaN
4	2
..	...
176	118
177	104
178	20236
179	Silesia (SL)
180	Notes: ^ Laboratory-confirmed cases, includi...

Warmia-Masuria (WN) \

0	NaN
1	NaN
2	1
3	1
4	NaN
..	...
176	51
177	22

```

178                                     857
179                               Warmia-Masuria (WN)
180 Notes:  ^ Laboratory-confirmed cases, includi...

                                     Greater Poland (WP) \
0                                         NaN
1                                         NaN
2                                         NaN
3                                         NaN
4                                         NaN
..                                       ...
176                                       92
177                                       92
178                                       5505
179                               Greater Poland (WP)
180 Notes:  ^ Laboratory-confirmed cases, includi...

                                     West Pomerania (ZP) \
0                                         NaN
1                                         NaN
2                                         2
3                                         NaN
4                                         NaN
..                                       ...
176                                       11
177                                       10
178                                       1096
179                               West Pomerania (ZP)
180 Notes:  ^ Laboratory-confirmed cases, includi...

                                     Poland daily \
0                                         1
1                                         0
2                                         4
3                                         1
4                                         5
..                                       ...
176                                       887
177                                       791
178                                       NaN
179                               Poland daily
180 Notes:  ^ Laboratory-confirmed cases, includi...

                                     Poland total \
0                                         1
1                                         1
2                                         5

```

```

3
4
..
176 64 689
177 65 480
178 65 480
179 Poland total
180 Notes: ^ Laboratory-confirmed cases, includi...

```

```

Sources
0 [714]
1 NaN
2 [715]
3 [17]
4 [19] [22]
..
176 [829]
177 [830]
178 NaN
179 Sources
180 Notes: ^ Laboratory-confirmed cases, includi...

```

[181 rows x 20 columns]

2.1 Uporządkowanie danych oraz przygotowanie wykresu.

```

[33]: lesser_pol_data = [load_csv["Date (CET)"], load_csv["Lesser Poland (MA)"]]
headers = ["Date", "Number of infections in Lesser Poland"]

df3 = pd.concat(lesser_pol_data, axis=1, keys=headers)

df4 = df3.drop([180, 179, 178])
df4

```

```

[33]:
Date Number of infections in Lesser Poland
0 4 March 2020 NaN
1 5 March 2020 NaN
2 6 March 2020 NaN
3 7 March 2020 NaN
4 8 March 2020 NaN
..
173 24 August 2020 61
174 25 August 2020 141
175 26 August 2020 118
176 27 August 2020 233
177 28 August 2020 139

```

[178 rows x 2 columns]

Get rid of NaNs

```
[34]: df5 = df4.replace(np.nan, 0)
df5
```

```
[34]:
```

	Date	Number of infections in Lesser Poland
0	4 March 2020	0
1	5 March 2020	0
2	6 March 2020	0
3	7 March 2020	0
4	8 March 2020	0
..
173	24 August 2020	61
174	25 August 2020	141
175	26 August 2020	118
176	27 August 2020	233
177	28 August 2020	139

[178 rows x 2 columns]

Convert date to numerical form.

```
[35]: date_elements = []
infected = []
for row in df5.itertuples(index=False):
    #print(row[0])
    date_elements.append(row[0].split())

# date elem has 178 records

for infections in df5.itertuples(index=False):
    infected.append(int(infections[1]))

t = np.linspace(0, len(date_elements), len(date_elements), endpoint=False)

# infected
```

```
[36]: import seaborn as sns
sns.set(style="darkgrid")
```

```
[37]: df5["Number of infections in Lesser Poland"].astype(int)
df5.info()
df5['Days'] = df5.index
```

```
<class 'pandas.core.frame.DataFrame'>
Int64Index: 178 entries, 0 to 177
```


Data columns (total 2 columns):

#	Column	Non-Null Count	Dtype
0	Date	178 non-null	object
1	Number of infections in Lesser Poland	178 non-null	object

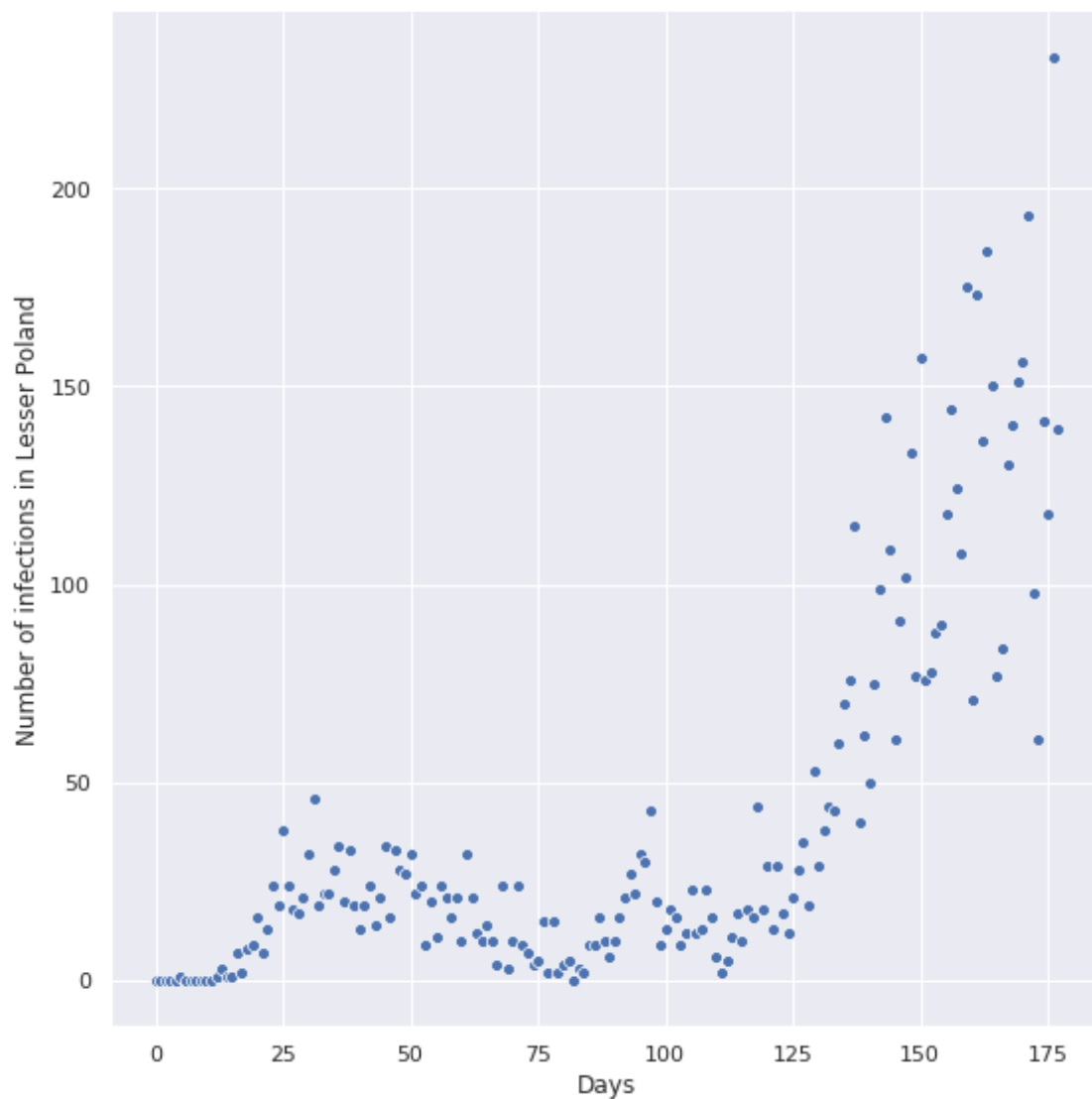
dtypes: object(2)

memory usage: 4.2+ KB

3 Wykres dni od wybuchu epidemii w województwie małopolskim.

```
[38]: sns.relplot(x="Days", y="Number of infections in Lesser Poland", data= df5, height=8)
```

```
[38]: <seaborn.axisgrid.FacetGrid at 0x7fdc57af0f10>
```

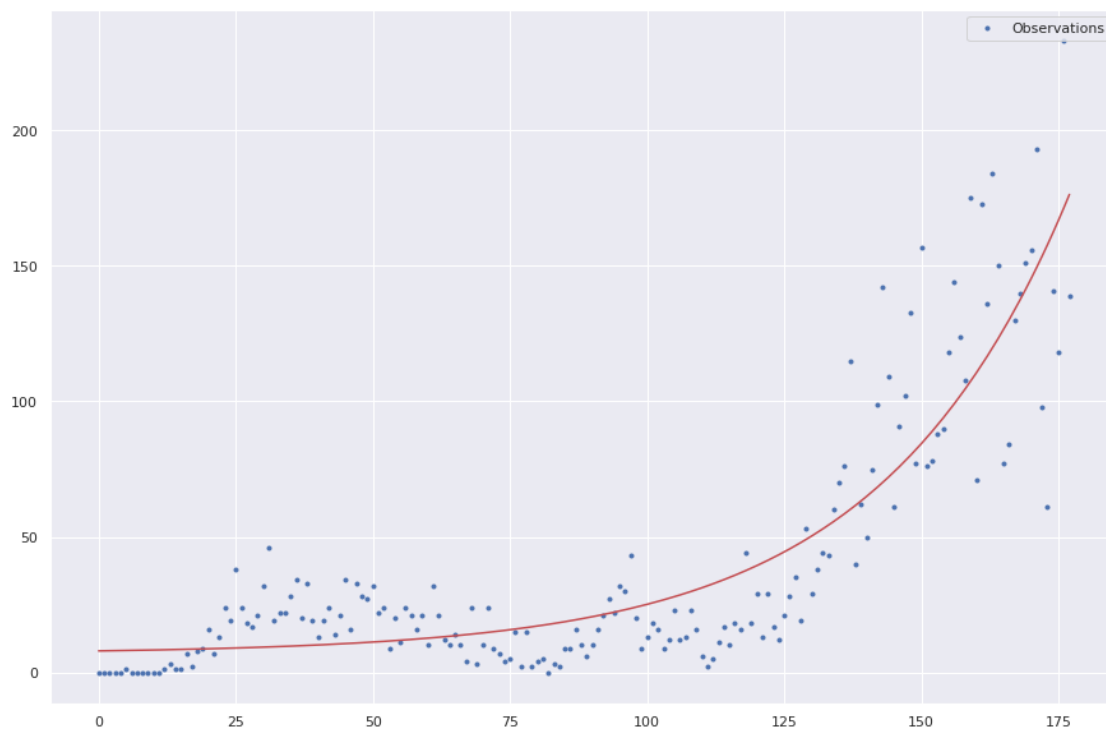


4 Model i funkcja fitująca tempo transmisji przypadków

```
[39]: def exp_fun(x, a, b, c, C):  
       return (a * np.exp((b * x) + c)) + C  
  
def fit_fun(X, y, fitted_fun, calls, plot=True):  
    popt, pcov = scipy.optimize.curve_fit(fitted_fun, X, y, maxfev=calls)  
    plt.figure(figsize=(15, 10))  
    plt.plot(X, y, ".", label="Observations")  
    y_fit = fitted_fun(X, *popt)  
    R2 = r2_score(y, y_fit)  
    print(f"Coefficient of determination: {R2}")  
    if plot:  
        plt.plot(X, y_fit, 'r-')  
        plt.legend()  
        plt.show()  
    #return popt, pcov, y_fit, R2
```

```
[40]: fit_fun(t, infected, exp_fun, 1000000)
```

Coefficient of determination: 0.7736392445688678



5 Wnioski

Z przeprowadzonej analizy wynika, że współczynnik poprawności estymacji wynosi $R^2 = 0.77$, daje nam to dostatecznie dobre przybliżenie rzeczywistości i oddaje dynamikę epidemii. Rozwój zakażeń wirusem SARS-COV2 w przybliżeniu oddaje funkcja wykładnicza. Autorzy w niedalekiej przyszłości użyją kolejnych modeli w celu porównania wyniku R^2 .

6 Model SIR

Model SIR to najprostszy model obrazowania epidemii, która mogłaby rozchodzić się w społeczeństwie. Nazwa pochodzi od pierwszych liter wyrazów oznaczających: S - Liczba podatnych na zakażenie (ang. Susceptible) I - Liczba zainfekowanych (ang. Infectious) R - Liczba osób, które przebyły chorobę lub zmarły (ang. Removed)

Założmy, że mamy populację składającą się z

$$N$$

osób, dla uproszczenia założmy, że populacja ta jest stała i liczba osób nie zmienia się w czasie, mamy równanie:

```
[41]: %%\latex
\begin{equation}
S + I + R = N
\end{equation}

Niech  $t$  oznacza zmienną niezależną oznaczającą jakąś chwilę czasu. Oznaczmy
→przez  $\beta$  współczynnik
zakaźności, a przez  $\gamma$  współczynnik czasu rekonwalescencji. Mamy wtedy
→zestaw następujących równań różniczkowych:
```

$$S + I + R = N \tag{1}$$

Niech t oznacza zmienną niezależną oznaczającą jakąś chwilę czasu. Oznaczmy przez β współczynnik zakaźności, a przez γ współczynnik czasu rekonwalescencji. Mamy wtedy zestaw następujących równań różniczkowych:

```
[42]: %%\latex
\begin{equation}
\frac{dS}{dt} = - \frac{\beta SI}{N}
\end{equation}
```

$$\frac{dS}{dt} = - \frac{\beta SI}{N} \tag{2}$$

```
[43]: %%latex
\begin{equation}
\frac{dI}{dt} = \frac{\beta SI}{N} - \gamma I
\end{equation}
```

$$\frac{dI}{dt} = \frac{\beta SI}{N} - \gamma I \quad (3)$$

```
[44]: %%latex
\begin{equation}
\frac{dR}{dt} = \gamma I
\end{equation}
```

$$\frac{dR}{dt} = \gamma I \quad (4)$$

Użyjemy biblioteki SciPy aby rozwiązać ten układ równań różniczkowych i zaprezentujemy rozwiązanie na wykresie:

```
[45]: from scipy.integrate import odeint

N = 1000
I_initial, R_initial = 1, 0 # One person is infected from the start
S_initial = N - I_initial - R_initial

beta, gamma = 0.3, 0.1

t = np.linspace(0, 250, 250)

def SIR_equations(y, t, N, beta, gamma):
    S, I, R = y
    Sprime = (- beta * S * I) / N
    Iprime = (beta * S * I / N) - gamma * I
    Rprime = gamma * I
    return Sprime, Iprime, Rprime

ivp = S_initial, I_initial, R_initial

solution = odeint(SIR_equations, ivp, t, args=(N, beta, gamma))
S, I, R = solution.T

fig = plt.figure()
ax = fig.add_subplot()
ax.plot(t, S / N, 'y', label='Podatni na zarażenie')
ax.plot(t, I / N, 'r', label='Zainfekowani')
```

```

ax.plot(t, R / N, 'g', label='Wyzdrowiałe jednostki')
ax.set_xlabel('Czas liczony w dniach')
ax.set_ylabel('Populacja liczona w tysiącach')
legend = ax.legend()
plt.show()

```



7 Komentarz

Z przedstawionych graficznie rozwiązań można prześledzić dynamikę rozwoju choroby zakaźnej (jak na przykład COVID-19). W zależności od współczynnika zakaźności/współczynnika rekonwalescencji. Krzywa żółta będzie miała znacznie szybszy spadek (co wynika po prostu z tego jak rozumie się pochodną - szybkość z jaką funkcja się zmienia). W momencie gdy nie będzie wystarczająco dużo jednostek zdolnych wejść w kontakt z patogenem, nastąpi wygaszenie choroby i co raz większa liczba jednostek trafi do kategorii wyzdrowiałych osób. Jako populacja mamy realny wpływ na to, żeby obniżyć β poprzez społeczną izolację, co pozwoli wypłaszczyć krzywą i w znacznym stopniu spowolnić rozwój wirusa oraz zapobiec paraliżowi służby zdrowia.

8 Referencje:

https://en.wikipedia.org/wiki/Template:COVID-19_pandemic_data/Poland_medical_cases