



Probabilidad Regresion

Materia:

Simulación

Docente:

Ing. Diego Quisi

Estudiante:

Ricardo Vinicio Jara Jara

Covid-19 infección en Ecuador. Modelos probabilísticos

Implementacion de un modelo probabilistico de infección por el virus Covid-19

Se realiza un análisis probabilistico simple del crecimiento de la infección en Python y el modelos para comprender mejor la evolución de la infección.

Se crea modelos de series temporales del número total de personas infectadas hasta la fecha (es decir, las personas realmente infectadas más las personas que han sido infectadas). Estos modelos tienen parámetros , que se estimarán por ajuste de probabilidad.

```
In [76]: # Importar las librerias para el analisis
import pandas as pd
import numpy as np
from datetime import datetime, timedelta
from sklearn.metrics import mean_squared_error
from scipy.optimize import curve_fit
from scipy.optimize import fsolve
from sklearn import linear_model
import matplotlib.pyplot as plt
from sklearn.linear_model import LinearRegression
from sklearn.preprocessing import PolynomialFeatures
%matplotlib inline

#Mis Datos COVID EC

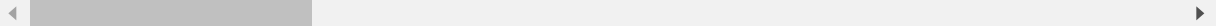
url = 'https://raw.githubusercontent.com/andrab/ecuacovid/master/datos_crudos/
ecuacovid.csv'

df = pd.read_csv(url)
df
```

Out[76]:

	muestras	muestras_pcr	muestras_pcr_nuevas	pruebas_rezagadas	muerteres_confirmadas	n
0	129	129	0	106	1	
1	206	206	77	178	2	
2	273	273	67	236	2	
3	354	354	81	296	2	
4	762	762	408	651	2	
...
251	629093	608521	5063	48617	8729	
252	632453	611881	3360	48275	8750	
253	636428	615856	3975	47265	8787	
254	641542	620970	5114	45942	8804	
255	643405	622833	1863	45756	8825	

256 rows × 32 columns



Imprimos los resultados y agregamos el numero del dia

```
In [10]: df = df.loc[:,['created_at','positivas_pcr']] #Selecciono las columnas de analisis

FMT = '%d/%m/%Y'
date = df['created_at']
df['created_at'] = date.map(lambda x : (datetime.strptime(x, FMT) - datetime.strptime("01/01/2020", FMT)).days)
x = list(df.iloc[:, 0]) # Fecha
y = list(df.iloc[:, 1]) # Numero de casos
df
```

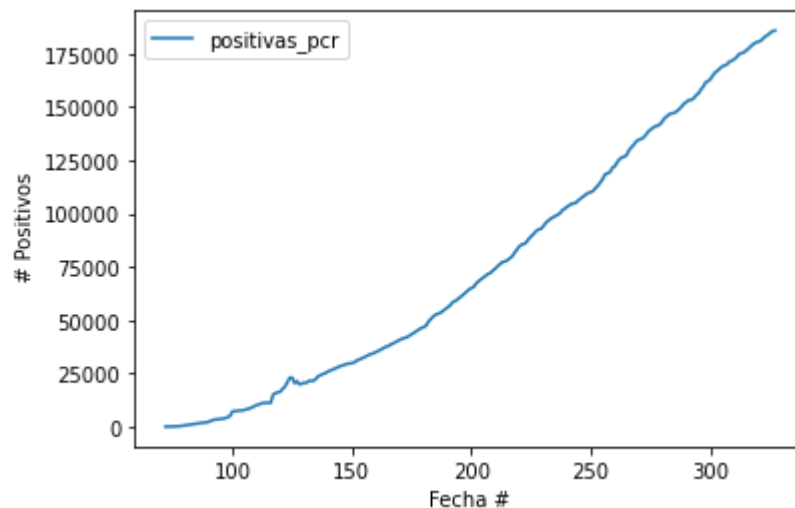
Out[10]:

	created_at	positivas_pcr
0	72	23
1	73	28
2	74	37
3	75	58
4	76	111
...
251	323	183246
252	324	183840
253	325	184876
254	326	185643
255	327	185944

256 rows × 2 columns

```
In [12]: ax = df.plot(x='created_at', y='positivas_pcr')
ax.set_xlabel("Fecha #")
ax.set_ylabel("# # Positivos")
```

Out[12]: Text(0, 0.5, ' # Positivos')



Ahora podemos analizar un modelo probabilístico para el examen.

El modelo basado en probabilidad

Para realizar una estimación del factor de crecimiento de los casos de Covid 19 en Ecuador calculamos la mediana, con esto obtenemos el valor medio de crecimiento de un conjunto de datos, con esto podemos obtener un factor de crecimiento o tasa de crecimiento de los nuevos casos.

```
In [184]: filtro = df["positivas_pcr"][61:] # Filtro los datos que se empezó a tener casos
#Obtenemos la mediana
media = filtro.mean()
mediana = filtro.median()
print("Mediana:", mediana)
print("Media:", media)
```

Mediana: 93985.0

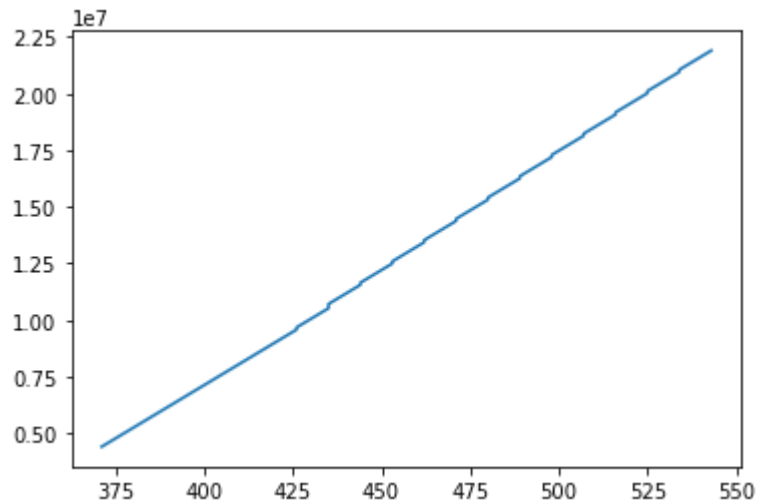
Media: 96812.89743589744

De la ecuación de la recta $y = mX + b$ nuestra pendiente «m» es el coeficiente y el término independiente «b»

```
In [186]: #Vamos a comprobar:
# según la media y la mediana podemos obtener la tasa de crecimiento y predecir su comportamiento.
# Cargamos los datos de total de casos
url = 'https://raw.githubusercontent.com/andrab/ecuacovid/master/datos_crudos/ecuacovid.csv'
df_t = pd.read_csv(url)
FMT = '%d/%m/%Y'
date = df_t['created_at']
df_t['created_at'] = date.map(lambda x : (datetime.strptime(x, FMT) - datetime.strptime("01/01/2020", FMT)).days)
df_t = df_t.loc[:, ['created_at', 'positivas_pcr']] #Selecciono las columnas de analisis
y = list(df_t.iloc[:, 1]) # Total casos
x = list(df_t.iloc[:, 0]) # Dias
#Realizamos un ejemplo de predicción
prediccion_siguiente = int(y[-1] + mediana)
print('Predicción: ', prediccion_siguiente)
```

Predicción: 279929

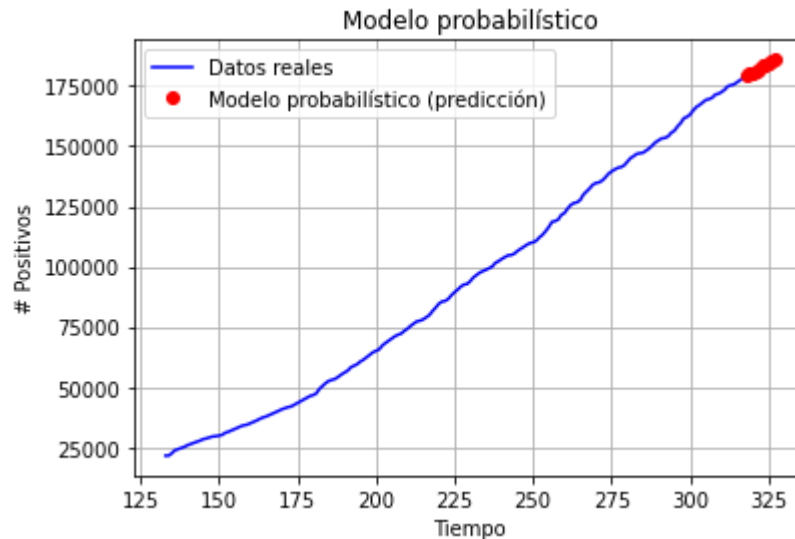
```
In [41]: # Quiero predecir cuántos "Casos" voy a obtener de aquí a 10 días.
for i in range(x[-1], x[-1]+10):
    x.append(i)
    y.append(int(y[-1] + mediana))
#Correlaciono con mis datos ya que desde día 300 comienza el crecimient
plt.plot(x[300:], y[300:])
plt.show()
```



Practica

1. Comparar el modelo de predicción matemático vs probabilidad.

```
In [187]: # Modelo probabilístico
ax = plt.figure().add_subplot(111, axisbelow=True)
ax.plot(x[61:len(x)-9], y[61:len(x)-9], 'b', label = 'Datos reales')
ax.set_title("Modelo probabilístico")
ax.set_xlabel('Tiempo')
ax.set_ylabel('# Positivos')
ax.plot(x[len(x)-10:], y[len(x)-10:], 'o', alpha=1, lw = 2, label = 'Modelo prob
abilístico (predicción)', color='red')
ax.legend()
ax.grid()
```



```
In [188]: #Prediccion de Mañanas
prediccion_siguiente = int(y[-1] + mediana)
print('Prediccion')
print('Mañana: ',prediccion_siguiente)
for i in range(x[-1], x[-1]+7):
    x.append(i)
    y.append(int(y[-1] + mediana))
print('Siguiente Semana:',y[-1]+7 )
```

Prediccion
Mañana: 279929
Siguiente Semana: 843846

```

In [218]: #Modelo Matematico
pf = PolynomialFeatures(degree = 4)
X = pf.fit_transform(np.array(x).reshape(-1, 1))

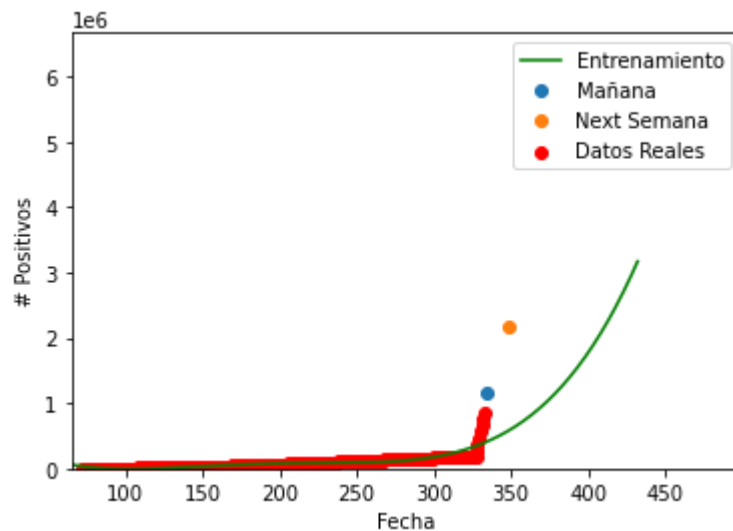
regresion_lineal = LinearRegression()

regresion_lineal.fit(X, y)
pred_x = list(range(0,max(x)+100))
puntos = pf.fit_transform(np.array(pred_x).reshape(-1, 1))
prediccion_entrenamiento = regresion_lineal.predict(puntos)
print ('Positivos Actuales:', int(prediccion_entrenamiento[-1]))

plt.plot(pred_x, prediccion_entrenamiento, color='green',label = 'Entrenamiento')
plt.scatter(x,y,label="Datos Reales",color="red")
plt.ylim((-300,max(y)*7.9))
plt.xlim((min(x)*0.9,max(x)*1.5))
plt.ylabel("# Positivos")
plt.xlabel("Fecha")
plt.plot(x[-1]+1,prediccion_entrenamiento[-1]-1, 'o', label = 'Mañana')
plt.plot(x[-1]+7,prediccion_entrenamiento[-1]+7, 'o', label = 'Next Semana')
plt.legend()
plt.show()

```

Positivos Actuales: 3168539



3. Retroceder un semana y comparar el modelo matematico vs probabilidad vs reales. Solo cargan los datos para generar los modelos menos 7 dias.

```

In [234]: semana_pasada=df['positivas_pcr'][-7:]
y_semanantes = list(df_t.iloc[:, 1]) # Total casos
x_semanantes = list(df_t.iloc[:, 0]) # Dias
for i in range(7):
    y_semanantes.pop(-1)
    x_semanantes.pop(-1)
#Mediana
print('Semana antes:', semana_pasada.median())

for i in range(x_semanantes[-1], x_semanantes[-1]+7):
    x_semanantes.append(i)
    y_semanantes.append(int(y_semanantes[-1]+mediana_semanantes))

for i in range(0,7):
    print('Dia: ',round(prediccion_semana_antes),' casos posibles')
    prediccion_semana_antes=round(semana_pasada + semana_pasada.median())
    break

plt.plot(x_semanantes[-30:], y_semanantes[-30:])
plt.show

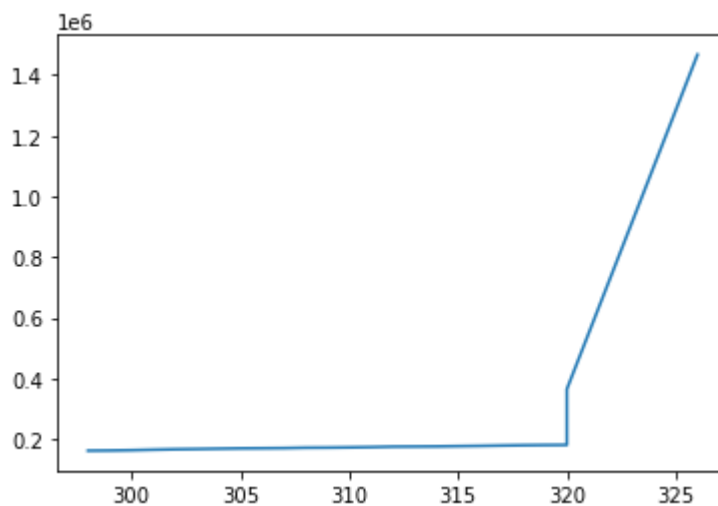
```

```

Semana antes: 183840.0
Dia: 249 364944.0
250 366090.0
251 367086.0
252 367680.0
253 368716.0
254 369483.0
255 369784.0
Name: positivas_pcr, dtype: float64 casos posibles

```

Out[234]: <function matplotlib.pyplot.show(*args, **kw)>



Analisis

Con un análisis previo en vista de los datos obtenidos se puede ver que el incremento de los casos es el doble o a veces están triplicando día a día. Por otro lado se ve un crecimiento exponencial aunque no tan prolongado, es decir no existe una concordancia en el crecimiento de los casos positivos por covid recuperación de los casos

Conclusiones

Puedo concluir que la simulación de los datos de regresión podemos evaluar a futuro eso sí me pareció interesante en la parte que la fecha de ser transformada a un número y cuando yo quiero predecir necesito definir este número mayor a los existentes. Sumo la fecha en el que esté pasando a número sumó una cierta cantidad y puedo ver qué datos obtendré en las siguientes fechas o en el siguiente número de días que yo quiero ver

Criterio personal

- politico
En el criterio económico la pandemia perjudico a todo tipo de empresa afectando más a las pequeños y medianos empresarios que recién estaba emprendiendo o que tenían pérdidas en su desarrollo de empresa por otro lado tampoco las personas naturales con ciertas necesidades económicas tuvieron el apoyo por parte del gobierno
- economico
En el criterio político se vio mucho que en Ecuador como no existe un claro gobierno que controlen aspectos en estado crítico. Más bien se ve que estas decisiones son tomadas por empresarios o empresas que compran el gobierno o en que dichas empresas son de funcionarios público. Por lo que se pudo ver claramente que las decisiones en envase de medicamentos para pacientes infectados fueron repartidos en los funcionarios públicos.
- social
En el aspecto el aspecto social se pudo ver que el país estaba dividido ya que gran parte era consciente y respetaba las leyes del confinamiento para evitando que el virus se propague. Por otra parte, existía ciudades del país que les hacían como un quemeimportismo a la situación no respetaban nuestro teniendo así que esta ciudad de sean controladas por las fuerzas públicas del país

Referencias

- https://www.researchgate.net/publication/340092755_Infeccion_del_Covid-19_en_Colombia_Una_comparacion_de_modelos_logisticos_y_exponenciales_aplicados_a_la_infeccion_por
(https://www.researchgate.net/publication/340092755_Infeccion_del_Covid-19_en_Colombia_Una_comparacion_de_modelos_logisticos_y_exponenciales_aplicados_a_la_infeccion_por)
- <https://www.aprendemachinelearning.com/regresion-lineal-en-espanol-con-python/>
(<https://www.aprendemachinelearning.com/regresion-lineal-en-espanol-con-python/>)