

ESTADÍSTICA

COLOQUIO FINAL

- Alumno: Gamba Francisco.
- Fecha de Entrega: 29/7/21.
- Tema: Casos de COVID-19.

Índice

DEFINICIÓN DE LOS DATOS.....	3
MUESTRA	3
FUENTE DE LA INFORMACIÓN	3
PROGRAMA	3
OBJETIVO.....	4
MUESTRA	5
DATOS	5
GRAFICO.....	5
OBSERVACIONES	5
CARACTERÍSTICAS	6
IMPORTANCIA	6
DETERMINACIÓN	6
• Media Muestral.....	6
• Mediana	6
• Varianza Muestral	6
• Desvío Estándar de la Muestra.....	6
• Coeficiente de Variabilidad	6
• Asimetría	6
• Curtosis.....	6
OBSERVACIONES	7
INFERENCIA	8
ESTIMACIÓN	8
DETERMINACIÓN	8
CONCLUSIÓN	9
SERIE CRONOLÓGICA	10
INTRODUCCIÓN	10
SERIE	10
SUAVIZADO	11
ÍNDICE ESTACIONAL	12
TENDENCIA.....	13
CÍCLICAS E IRREGULARES	14
DATOS	15

<u>APORTE PERSONAL</u>	<u>16</u>
FIABILIDAD DE LOS DATOS	16
VISUALIZACIÓN DE LAS OLAS	16
TENDENCIA CRECIENTE	16
EXPECTATIVAS A FUTURO	16
<u>BIBLIOGRAFÍA.....</u>	<u>17</u>

Definición de los datos

Muestra

La muestra a estudiar será la cantidad de Casos COVID19 diarios en la república argentina, es decir la cantidad de casos nuevos por día que se produjeron en el territorio argentino.

$$X = \text{Casos nuevos en un día}$$

Fuente de la información

Estos datos son públicos por obligación del estado y se encuentra en la web del ministerio de salud siguiente.



Ministerio de Salud
Argentina

[Haga clic para Ir a la web.](#)

Programa

Se realizó un programa el cual se encargó de armar todas las muestras a partir del archivo de casos covid 19 brindado por el gobierno.

```
-----CASOS POSITIVOS COVID 19-----  
  
-----MENU-----  
  
1 - Instrucciones  
2 - Poblacion  
3 - Muestreo  
ESC - Salir  
  
-----Francisco Matias Gamba-----
```

Menú Principal del programa.

Este programa realiza muestras de las mediciones agrupando por zona, fecha inicial y fecha final pudiendo así obtener una muestra para los casos positivos en una provincia o el país dentro de un periodo.

Además de armar la muestra el programa nos permite consultar las características de la muestra cargada en memoria, consultar por un intervalo de confianza para la media de dicha muestra y armar una serie de datos agrupando en meses.

```
-----MUESTREO-----  
-----MENU-----  
1 - Ingresar Parametros  
2 - Ver Muestra  
3 - Caracteristicas  
4 - Inferencia  
5 - Serie  
6 - Guardar  
ESC - Salir  
-----
```

Menú de las opciones de muestreo.

La última opción de muestreo nos permite guardar la muestra en un archivo .csv el cual es compatible con Excel. Este archivo contiene en orden lo siguiente:

- Descripción de la muestra.
- Características de la muestra.
- Estimación para la esperanza con una confianza del 95%.
- Muestra en sí misma.
- Serie cronológica discretizada en meses.

Para finalizar con lo correspondiente al programa dejamos un enlace a Google Drive para la descarga del mismo.



[Descargar programa Casos COVID 19.exe](#)

Objetivo

El objetivo de este trabajo fue determinar conclusiones sobre la muestra en base a la estadística determinando:

- Determinación de características de la muestra.
- Estimación del parámetro media de casos covid 19.
- Análisis de la serie de casos covid discretizada en meses.
- Visualizar la tendencia de los casos.

Muestra

Datos

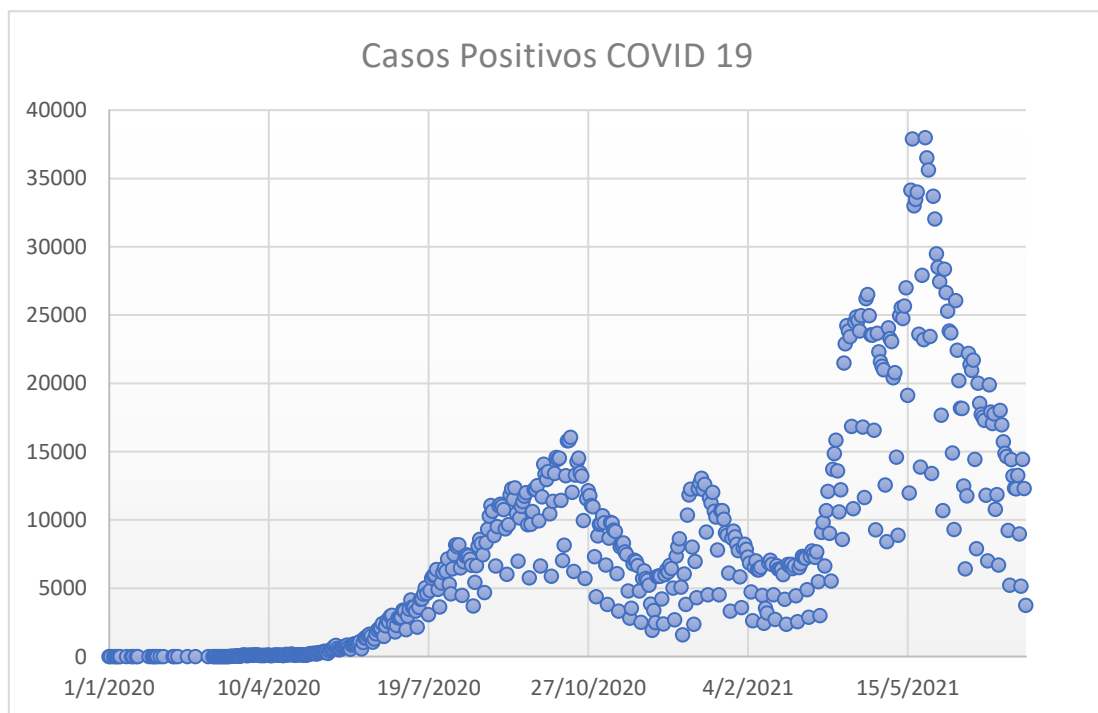
Los datos de todos los casos nuevos de COVID 19 en la Argentina son los siguiente.



[Ir a enlace para visualizar el documento .xlsx](#)

Grafico

El gráfico de la muestra es el siguiente.



Observaciones

Podemos observar la famosa curva de casos Covid-19 donde en la Argentina vivimos una primer ola durante el 2020 y la llamada segunda ola para el 2021 con una notoria creciente de la segunda respecto a la primera, también se puede observar una notable bajada de la segunda ola probablemente debido a la aplicación de la vacuna en gran parte de la población para las fechas más actuales registradas, otra probable causa del aplanamiento puede ser debido a la inmunidad de rebaño que inmunizo a la población contagiada anteriormente reduciendo la cantidad de casos nuevos.

Características

Importancia

Para detallar en profundidad la muestra determinamos las características de esta, además estas luego nos sirven para determinar un intervalo de confianza para la estimación del parámetro de casos esperados.

Determinación

Se determinaron las siguientes características de la muestra a través del Software Casos Covid 19.

- Media Muestral

$$\bar{x} = \frac{\sum_{i=0}^n x_i}{n} = 8395.14$$

- Mediana

$$X_m \rightarrow P(x < X_m) = 0.5 \rightarrow X_m = 6702$$

- Varianza Muestral

$$S^2 = \frac{\sum_{i=0}^n (x_i - \bar{x})^2}{n - 1} = 63115000$$

- Desvió Estándar de la Muestra

$$S = \sqrt{S^2} = 7944.49$$

- Coeficiente de Variabilidad

$$Cv = \frac{S}{\bar{x}} = 0.94632$$

- Asimetría

$$As = \frac{\sum_{i=0}^n (x_i - \bar{x})^3}{S^3} = 1.23938$$

- Curtosis

$$C = \frac{\sum_{i=0}^n (x_i - \bar{x})^4}{S^4} = 4.3233$$

Observaciones

En principio podemos ver que durante la pandemia hasta el día de hoy se dio un promedio de 8107 casos nuevos por día, con una variabilidad de 95%, pronto determinaremos el valor esperado real de casos promedio.

Otro punto que podemos observar que la muestra tiene una asimetría positiva, lo que como bien sabemos nos indica que los valores mas cercanos a la fecha actual tienen mayor peso, esto se puede observar en la grafica del lado de la segunda ola que es muy superior a la primera.

Inferencia

Estimación

La variable que estimaremos aquí será la cantidad de casos nuevos esperados por día para la republica argentina.

$$\text{Variable a estudiar} \rightarrow \mu$$

Con la idea de determinar este valor promedio como valor esperado a lo largo de todo el ciclo de contagio, es decir desde el primer día en que se registro un caso covid con regularidad, utilizamos los datos con rango desde el 1 de enero del 2020 hasta el día actual 28 de Julio del 2021.

$$\text{Rango de datos} \rightarrow 01/01/2020 \text{ hasta } 28/07/2021$$

Luego tomando a la media muestral como estimador puntual de la esperanza y al valor del desvío estándar de la muestra como estimador puntual del desvío.

$$\bar{x} \rightarrow \text{Estimador puntual} \rightarrow \mu$$

$$S \rightarrow \text{Estimador puntual} \rightarrow \sigma$$

Como la muestra es mayor a 30 días podemos utilizar el modelo normal para la estimación de la esperanza.

$$n > 30 \rightarrow \bar{x} \sim N\left(\mu, \frac{\sigma}{\sqrt{n}}\right)$$

Para finalmente determinar el intervalo de confianza.

$$\mu \in \left(\bar{x} \pm Z_{1-\frac{\alpha}{2}} \frac{S}{\sqrt{n}}\right)$$

Donde $1 - \frac{\alpha}{2}$ es la confianza para el intervalo.

Determinación

Con una confianza del 95% el intervalo determinado para la cantidad de casos esperada en la Argentina es el siguiente:

$$\mu \in (8395.14 \pm 673.19)$$

Conclusión

Dicho esto, podemos afirmar con un 95% de confianza que el valor esperado para la cantidad de casos covid nuevos diarios está entre los 8395 casos y 674 casos.

En Argentina el valor esperado de casos esta entre 8395 y 674 casos

Serie Cronológica

Introducción

Se discretizo toda la muestra de casos covid 19 por meses significativos desde el comienzo de la expansión viral en la Argentina.

$$Y_i = \text{Cantidad de casos en Argentina para el mes } i$$

Existen registros de casos en años anteriores al 2019 en la muestra, quizás por error de la muestra o casos reales de otra cepa el caso es que analizaremos la serie descrita desde enero del 2020 hasta Julio 2021 con el error de estos últimos 3 días faltantes de Julio.

$$\text{Rango de Datos} \rightarrow \text{Enero 2020 a Julio 2021}$$

Por ser una cantidad en números enteros la variable será una variable discreta, por lo que nuestros resultados deberán ser números enteros.

Los datos serán tratados como series anuales discriminadas mensualmente para los años 2020 y 2021 a través del modelo de hipótesis multiplicativa.

$$Y_i = T_i \cdot S_i \cdot C_i \cdot I_i$$

Luego determinaremos a través de un suavizado exponencial y un promedio entre estos para cada mes un índice de estacionalidad que nos sirve para desestacionalizar los datos.

$$Y_{di} = \frac{Y_i}{IE}$$

Es con estos datos que podremos conseguir a través de una regresión lineal la tendencia de la serie la cual nos servirá para realizar estimaciones sobre los casos en la argentina.

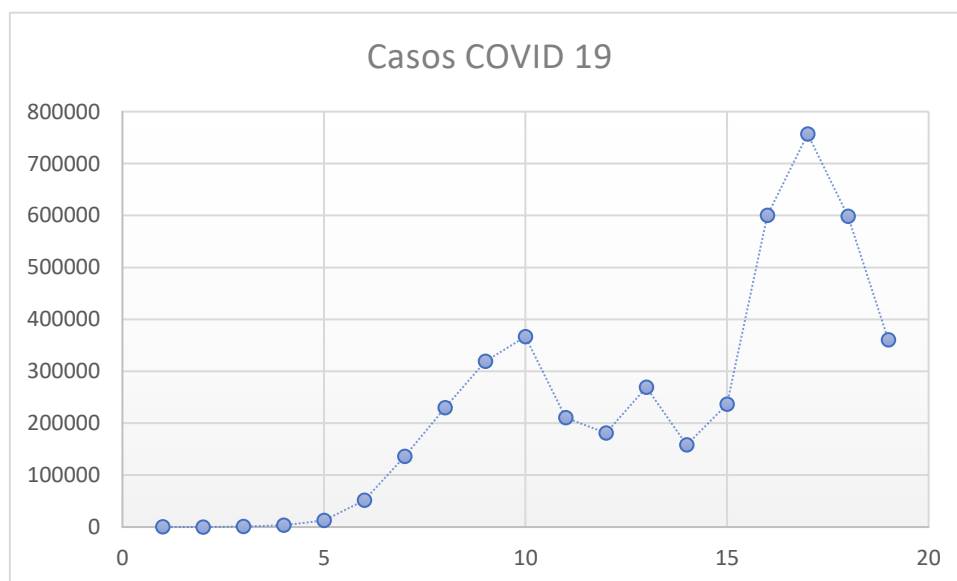
$$\text{Regresión Lineal de } Y_{di} \rightarrow \text{Tendencia } T_i$$

Serie

La tabla siguiente muestra los valores de la serie determinados por el programa.

Ano	Mes	Positivos
2020	Enero	65
2020	Febrero	8
2020	Marzo	1088
2020	Abril	3443
2020	Mayo	12931
2020	Junio	51317
2020	Julio	135827
2020	Agosto	229584
2020	Septiembre	319098
2020	Octubre	366255
2020	Noviembre	210710
2020	Diciembre	181151
2021	Enero	269163
2021	Febrero	158328
2021	Marzo	236394
2021	Abril	600223
2021	Mayo	756911
2021	Junio	598443
2021	Julio	360463

El siguiente gráfico describe los valores de la tabla.

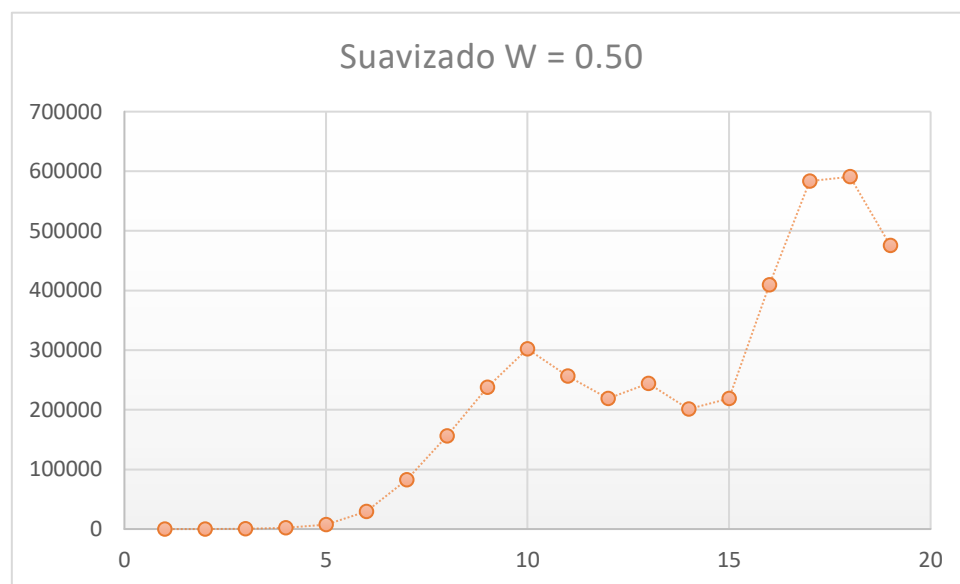


Suavizado

Se eliminó lo más que se pudo la irregularidad de los datos a través de un suavizado exponencial, utilizando un $w=0.5$.

Ano	Mes	W = 0.50
2020	Enero	65
2020	Febrero	36.5
2020	Marzo	562.25
2020	Abril	2002.625
2020	Mayo	7466.8125
2020	Junio	29391.9063
2020	Julio	82609.4531
2020	Agosto	156096.727
2020	Septiembre	237597.363
2020	Octubre	301926.182
2020	Noviembre	256318.091
2020	Diciembre	218734.545
2021	Enero	243948.773
2021	Febrero	201138.386
2021	Marzo	218766.193
2021	Abril	409494.597
2021	Mayo	583202.798
2021	Junio	590822.899
2021	Julio	475642.95

Vemos que el grafico se vio suavizado en la siguiente figura.

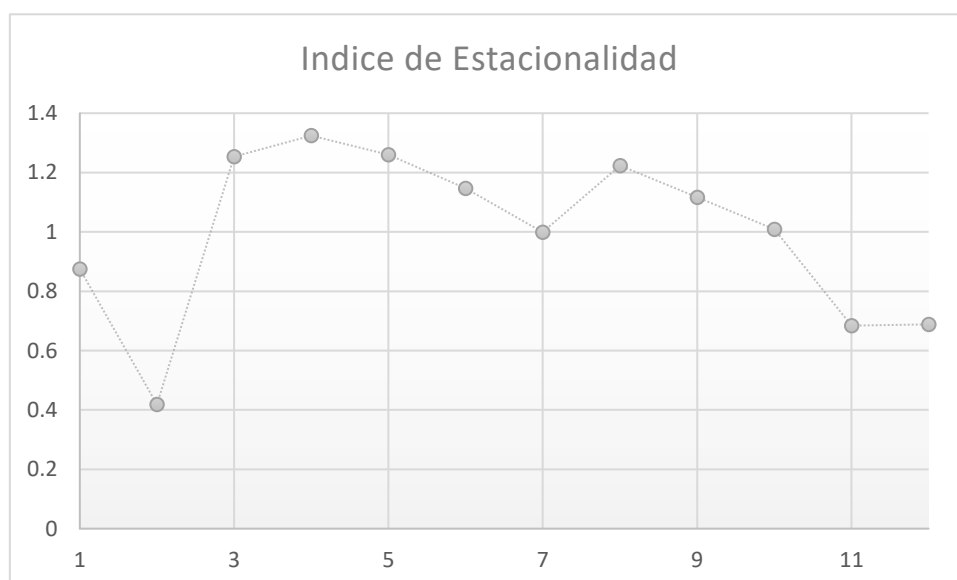


Índice Estacional

Luego se determinó el índice estacional para cada mes, a causa de la falta de repetitividad del ciclo anual en los casos de Covid19 algunos índices contaron con un único valor lo que no nos dejó mucha elección que determinar, para los casos con dos registros se tomó el valor promedio.

Mes/Año	2020	2021	Promedio	IE
Enero📅	1	1.1033587	1.05167935	0.87472515
Febrero📅	0.21917808	0.78715954	0.50316881	0.41850628
Marzo📅	1.93508226	1.08057829	1.50783028	1.25412471
Abril📅	1.71924349	1.46576537	1.59250443	1.3245517
Mayo📅	1.73179653	1.29785214	1.51482434	1.25994196
Junio📅	1.74595685	1.01289744	1.37942714	1.14732652
Julio📅	1.6442065	0.75784367	1.20102509	0.99894217
Agosto📅	1.47078036		1.47078036	1.22330877
Septiembre	1.34301995		1.34301995	1.11704516
Octubre📅	1.21306141		1.21306141	1.00895327
Noviembre	0.82206449		0.82206449	0.68374498
Diciembre	0.82817737		0.82817737	0.68882932
		Total	14.427563	12
		FC	0.8317413	

Los índices estacionales determinados se pueden observar en el siguiente gráfico.



Del índice podemos ver que para los meses mas cálidos hay un descenso de casos, pero solo para la estación del verano esto podría ser causa de que en dichas épocas el virus sea menos resistente en el exterior debido a las altas temperaturas, otra causa podría deberse a la mayor fortaleza física de la población debido a que las estaciones mas invernales suelen someter a una persona físicamente con mayor fuerza.

Tendencia

Para la tendencia, luego de determinar los valores del índice estacional se desestacionalizo la serie como antes se explicó, a esta nueva serie de datos se le determino una regresión lineal a través del método de mínimos cuadrados el cual nos determinó la siguiente expresión.

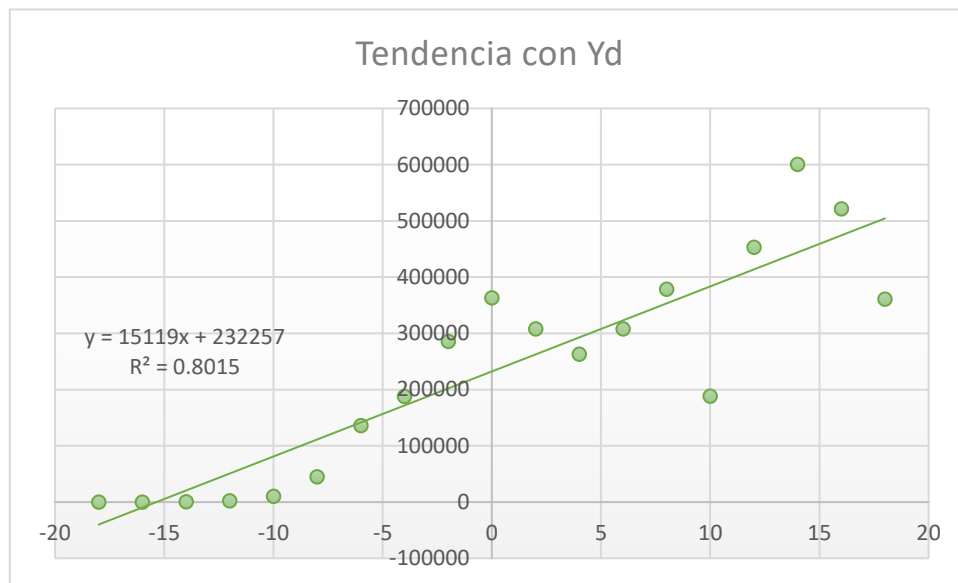
$$T = 15119x + 232257$$

De despejar del coeficiente de determinación obtenemos el coeficiente de correlación siguiente.

$$R = 0.90$$

Este coeficiente nos habla de lo buena que es el ajuste como regresión lineal, al estar tan próximo a 1.

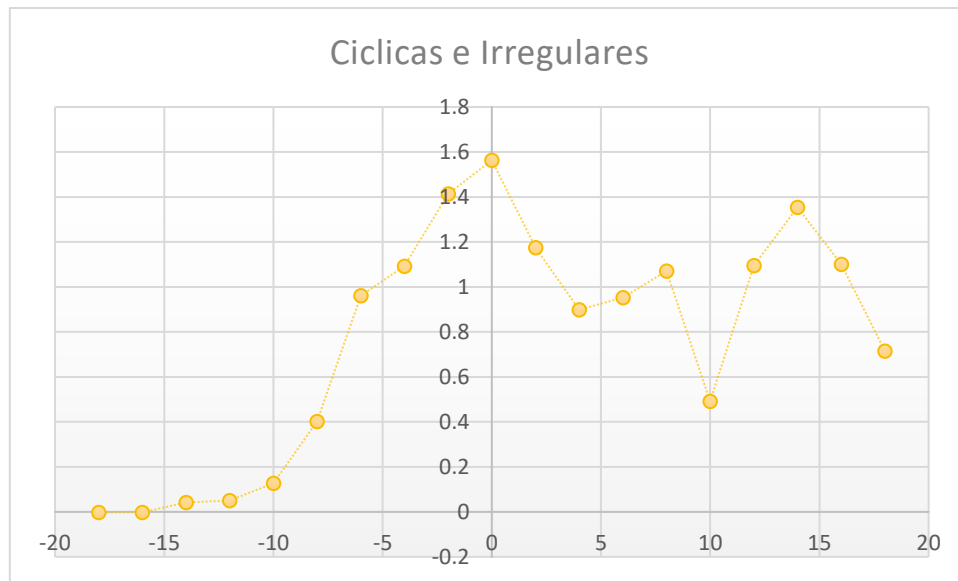
Veamos como ajusta gráficamente la tendencia.



Observemos como la tendencia para los primeros valores nos arroja valores no lógicos, esto se da porque los casos de Covid19 para valores iniciales no son significativos para el resto de la serie recordemos que la curva de Covid19 al principio se comporta de manera exponencial aumentan los casos de forma desmedida para las primeras mediciones lo que es de esperarse que en los tramos iniciales la curva no tenga un comportamiento lineal, pero observemos que el crecimiento de casos en general si tuvo un comportamiento creciente lineal como muestra la línea de tendencia.

Cíclicas e Irregulares

Despejando de la serie desestacionalizada la tendencia obtuvimos las cíclicas e irregulares.



Datos

La serie determinada por el software y luego tratada para poder realizar el análisis de la serie se encuentra disponible en el siguiente enlace.



[Ir a enlace para visualizar el documento .xlsx](#)

Aporte Personal

Fiabilidad de los datos

La fuente de los datos proviene del gobierno, los valores coinciden al menos en proporciones de crecimiento respecto a los valores de otras fuentes no gubernamentales lo que da fiabilidad en los datos, eso si hay un porcentaje menor de casos para las mediciones brindadas por el estado respecto a las brindadas por fuentes internacionales como Github o Google.

Visualización de las olas

El grafico de los casos diarios nos permitió observar como los casos de covid se manifestaron en forma de dos olas a lo largo del tiempo en la Argentina la primera un tanto mas suave que la segunda que impacto violentamente a la población dejando los peores números para la nación.

Tendencia Creciente

Se puede observar en el análisis de la tendencia que los casos tienen una tendencia creciente, esto es problemático, de seguir así los casos seguir aumentando con el paso del tiempo, por suerte para los datos más próximos se nota un decrecimiento probablemente debido a la campaña de vacunación que últimamente se esta haciendo mas notoria, esperemos que esto se vea impactado en la tendencia para las próximas mediciones.

Expectativas a futuro

Los datos nos muestran que la segunda ola esta decayendo al punto de que nos encontramos en los valores mas bajos de la primera, si esta tendencia reciente se mantiene puede que reforzada por la campaña de vacunación y el cambio estacional con la venida de los meses más cálidos la curva decaiga a niveles inferiores a los que nos viene acostumbrando durante el ultimo año.

Bibliografía

Todas las fuentes citadas para la realización de este trabajo.

- [Raymond H. Myers y Ronald E. Walpole \(2007\) - Probabilidad y estadística para ingenieros – Pearson Educación. ISBN 9702609364.](#)
- [Correlation Coefficient - Bivariate Normal Distribution.](#)
- [Mann, Prem S. \(1995\). Introductory Statistics \(2nd ed.\). Wiley. ISBN 0-471-31009-3.](#)
- [John Francis Kenney \(1946\) - Mathematics of Statistics, Volumen 2 - Van Nostrand, 1946.](#)
- [Inferencia Estadística \(2ª Edición Revisada\). ISBN 978-84-9828-131-6.](#)