Estimación de la Tasa de Letalidad

durante el curso de una epidemia con una aplicación a Covid-19 en Argentina

- Agustín Alvarez -
Coautores: Marina Fragalá y Marina Valdora

XVII Congreso Monteiro - Junio 2023

¿Cómo se suele entender la tasa de letalidad?

En pandemias como el Covid-19, nos interesa la tasa de letalidad sobre infectados (**ifr**), pero obtenerla tiene varias dificultades:

- Conocer el total de infectados. Hay muchos asintomáticos.
- Defasaje temporal entre contagio y muerte (que produce dicha enfermedad)
- Demoras en carga de datos
- Otras.

Tasa de letalidad sobre confirmados

- Una opción es concentrarse en la tasa de letalidad sobre confirmados (**cfr**) para no tener la dificultad de no conocer los infectados.

- Para hablar de tasa de letalidad hay que definir un período de tiempo (de contagio) de interés y una población.

- La tasa de letalidad (**cfr**) será la **probabilidad de fallecer** entre los que se confirman en dicho período.

¿Cómo se suele informar la tasa de letalidad?

El cálculo ***naive*** que se reporta el día t, $CFR_N(t)$, se calcula como:

$$\text{proporción de muertos hasta el día } t = \frac{\text{fallecidos hasta el día } t}{\text{confirmados hasta el día } t}.$$

Intenta dar una **Letalidad acumulada sobre confirmados.**

¿Qué error tiene este simple cálculo?

Muchos de las personas confirmadas en los últimos días morirán, pero aún no lo hicieron. Por lo tanto, el cálculo ***naive*** subestima la probabilidad que deseamos estimar.

Mejorar esta estimación es nuestro problema de interés.

Primeras definiciones

Notamos (Modelamos)

- p_d : probabilidad de morir para casos confirmados durante el día d.
- c_d : cantidad de confirmados durante el día d.
- $M_{d,i}$: variable aleatoria que toma valor 1 si el i-ésimo confirmado del día d muere por Covid y 0 en caso contrario. Asumimos $\{M_{d,i}\}_{d,i}$ independientes. $M_{d,i} \sim \mathrm{Be}(p_d)$
- $M_{d,i}(t)$: variable aleatoria Bernoulli que toma valor 1 si el i-ésimo confirmado del día d muere el día t o antes. Asumimos $\{M_{d,i}(t)\}_{d,i}$ independientes para cada t fijo.

Llamamos:

$$M_{t,f} = \sum_{d=0}^{t} \sum_{i=1}^{c_d} M_{d,i} \quad \text{y} \quad M(t) = \sum_{d=0}^{t} \sum_{i=1}^{c_d} M_{d,i}(t).$$

¿Cuándo se pueden calcular?

Observar que:

- $M_{t,f} = \sum_{d=0}^{t} \sum_{i=1}^{c_d} M_{d,i}$ es la cantidad que termina muriendo por Covid entre los confirmados hasta el día t. Se puede calcular recién cuando fallecieron los que iban a fallecer entre los casos confirmados hasta el día t

- $M(t) = \sum_{d=0}^{t} \sum_{i=1}^{c_d} M_{d,i}(t)$ es la cantidad de falleció hasta el día t entre los casos confirmados hasta el día t. Se puede calcular el mismo día t si no hay demora en la carga de datos

La estimación ***naive*** que se suele reportar el día t para informar la tasa de letalidad, está dada por

$$CFR_N(t) = \frac{M(t)}{\sum\limits_{d=0}^{t} c_d},$$

que resulta menor o igual a la que desearíamos calcular

$$CFR_F(t) = \frac{M_{t,f}}{\sum\limits_{d=0}^{t} c_d},$$

pero así definida no puede ser computada el día t.

Case Fatality Rate

Definición

Llamamos **Tasa de Letalidad** hasta el día t a

$$cfr(t) = \mathbb{E}\left(CFR_F(t)\right).$$

Observaciones

$$cfr(t) = \sum_{d=0}^{t} c_d p_d \Big/ \sum_{d=0}^{t} c_d = \sum_{d=0}^{t} \omega_d p_d.$$

1. Es una suma ponderada de las probabilidades diarias p_d.
2. Si $p_d = p$ entonces $cfr(t) = p$.
3. Interpetación posible de $cfr(t)$: probabilidad de que una persona confirmada hasta el día t, muera.
4. **Objetivo inicial**: Poder estimar $cfr(t)$ el mismo día t con un estimador insesgado.

- La letalidad sobre personas confirmadas p_d puede variar en el tiempo.
  - Cambiando la definición de caso sospechoso, o de caso confirmado.
  - Durante un proceso de vacunación.
  - Aparición de algún tratamiento.
  - Cambia la conformación de confirmados.
  - etc.

Propuesta (Idea)

Supongamos que

- entre las personas que se contagiaron hace justo 5 días, sabemos que hasta hoy fallecieron 6.
- entre quienes mueren por COVID, la probabilidad de morir en 5 días o menos es $3/10$,
- ¿Cómo estimamos $x=$ cantidad de gente que va a morir entre los que se confirmaron hace 5 días

Planteamos $x \cdot 3/10 = 6$ y despejamos $x = \frac{6}{3/10} = 20$.

- Sumando esas estimaciones para todos los que se fueron confimando de COVID desde el día 0 hasta hoy (día t), podemos estimar cuántos van a terminar muriendo de los confirmados hasta hoy.

Propuesta

Consideramos

1. la variable aleatoria $T_{d,i}$ cuenta cantidad de días desde el diagnóstico hasta la muerte del i-ésimo confirmado del día d.
2. Suponemos que la v.a. $T_{d,i}|\{M_{d,i}=1\}$: tiene distribución acumulada F_d .

Dado que:

$$P(M_{d,i}(t)=1)=P(M_{d,i}=1)P(T_{d,i}\leq t-d|M_{d,i}=1)=p_dF_d(t-d)$$

resulta

$$M_{d,i}(t)\sim \text{Bernoulli}(p_dF_d(t-d)).$$

- Sea $M_d(t) = \sum_{i=1}^{c_d} M_{d,i}(t)$ la cantidad de personas fallecidas hasta el día t entre los casos confirmados durante el día d $(d \leq t)$.
- Observar que $\mathbb{E}(M_d(t)) = c_d p_d F_d(t-d)$. Luego $\mathbb{E}\left(\frac{M_d(t)}{F_d(t-d)}\right) = c_d p_d$
- Recordar que $cfr(t) = \frac{\sum_{d=0}^{t} c_d p_d}{\sum_{d=0}^{t} c_d}$.
- Proponemos como estimador $CFR(t) = \frac{\sum_{d=0}^{t} \frac{M_d(t)}{F_d(t-d)}}{\sum_{d=0}^{t} c_d}$
- Observar que $\mathbb{E}(CFR(t)) = cfr(t)$

Propuesta

Definición

Sea $Z_{d,i}(t) = \frac{M_{d,i}(t)}{F_d(t-d)}$ se puede expresar

$$CFR(t) = \frac{\sum_{d=0}^{t} \sum_{i=1}^{c_d} Z_{d,i}(t)}{\sum_{d=0}^{t} c_d}.$$

Otra forma de ver $CFR(t)$

$CFR(t)$ es la media muestral de las $Z_{d,i}(t)$

- donde las $Z_{d,i}(t)$ son independientes para cada t fijo.

Encontramos un Teorema Central del Límite para arreglos triangulares.

Teorema

Supongamos que para cada $t \in \mathbb{N}$, $X_{t,1}, X_{t,2}, \ldots, X_{t,r_t}$ son variables aleatorias independientes. Sea $S_t = X_{t,1} + X_{t,2} + \ldots + X_{t,r_t}$. Supongamos que $\mathbb{E}(X_{t,k}) = 0$ para todo t y k, y que las varianzas $\mathbb{E}(X_{t,k}^2)$ son finitas. Llamamos $\sigma_t^2 = \mathbb{V}(S_t)$. Si la condición de Lindemberg se satisface, es decir:

$$\lim_{t \to \infty} \frac{1}{\sigma_t^2} \sum_{k=1}^{r_t} \mathbb{E}\left[X_{t,k}^2 \cdot \mathbf{1}_{\{|X_{t,k}| > \varepsilon \sigma_t\}}\right] = 0$$

para todo $\varepsilon > 0$, donde $\mathbf{1}_{\{\ldots\}}$ es la función indicadora,luego

$$S_t / \sigma_t \xrightarrow{D} N(0,1)$$

Consistencia y distribución asintótica

Teorema

Para cada $t \in \mathbb{N}$, sea la sucesión de variables $M_{d,i}(t)$, $0 \leq d \leq t$, $1 \leq i \leq c_d$ v.a.independientes Bernoulli$(p_d F_d(t-d))$.
Consideramos

- A1: $\inf_d F_d(0) = D > 0$.
- A2: $\inf_d p_d = I > 0$.
- A3: $\sup_d p_d = S < 1$.

Asumiendo que el número total de casos confirmados hasta el día t, $r_t = \sum_{d=0}^{t} c_d \xrightarrow{t\to\infty} \infty$, luego

1. si vale A1, $CFR(t) - cfr(t) \xrightarrow{p} 0$
2. si valen A1 hasta A3, entonces $\frac{CFR(t)-cfr(t)}{\sqrt{\mathbb{V}(CFR(t))}} \xrightarrow{D} N(0,1)$.

donde

$$\mathbb{V}(CFR(t)) = \frac{\sum_{d=0}^{t} c_d p_d (1 - p_d F_d(t-d))/F_d(t-d)}{r_t^2}.$$

Estimación por Intervalos de Confianza

A partir del Teorema anterior derivamos los intervalos de confianza asintóticos para $cfr(t)$

$$CFR(t) \pm z_{1-\alpha/2}\sqrt{\mathbb{V}(CFR(t))}.$$

Para el cálculo de dichos intervalos utilizaremos estimaciones de F_d y p_d. Para cada $d^* \leq t$ proponemos estimar p_d^* como

$$\hat{p}_{d^*} = \frac{\sum_{d=d^*-3}^{d^*+3} \sum_{i=1}^{c_d} Z_{d,i}(t)}{\sum_{d=d^*-3}^{d^*+3} c_d}.$$

Otros estimadores de $cfr(t)$

Garske el al. (2009) definieron un estimador de la Tasa de Letalidad utilizando F constante con respecto a d como

$$CFR_G(t) = \frac{M(t)}{\sum_{d=0}^{t} c_d F(t-d)}.$$

Si lo modificamos reemplazando F por F_d obtenemos $CFR_G^M(t)$. Acá

$$\mathbb{E}(CFR_G^M(t)) = \frac{\sum_{d=0}^{t} c_d F_d(t-d) p_d}{\sum_{d=0}^{t} c_d F_d(t-d)} = \sum_{d=0}^{t} \omega_d^* p_d.$$

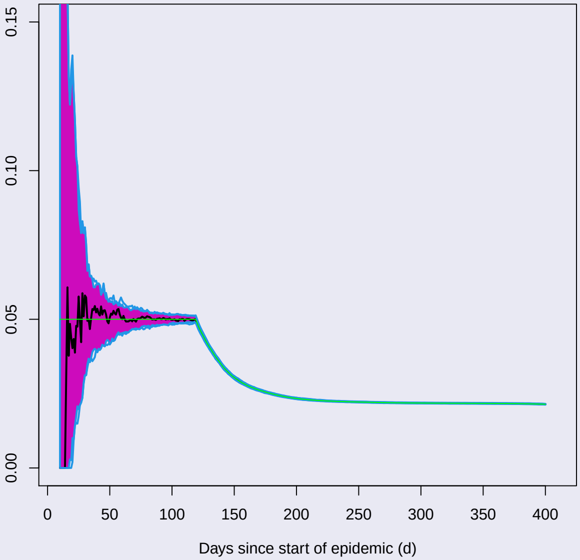
No podemos asegurar insesgadez con respecto a $cfr(t)$ salvo que $p_d = p$.

Estudio de Monte Carlo

- Planteamos escenarios usando las cantidades de confirmados diarios de Argentina y de India, de los primeros 400 días. 2 posiblidades para los c_d
- Para los p_d tenemos 2 posibilidades: usamos los p_d ***abruptos*** y los empíricos "suavizados" de Argentina.
- Los p_d abruptos tienen la característica de que $p_d = 0.05$ desde del día 0 hasta el día 120 y; a partir del día 121 hasta el día 400 se mantiene constante $p_d = 0.02$.
- Planteamos para generar las simulaciones: $F_d = F$ para todo d, donde F es una Binomial Negativa con ceros inflados (ZINB). 2 modelos para la F: la que mejor ajusta en Argentina u otra ZINB con esperanza a la mitad (mueren más rápido).
- Para estimar: usamos la F que generó los datos (F conocida) o estimamos la F.

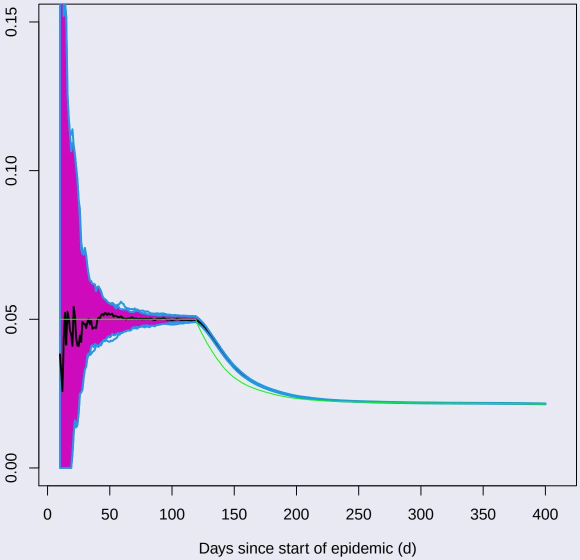
Boxplot funcional del Estimador Nuestro

Letalidad abrupta, c_d de India, sin estimar F.



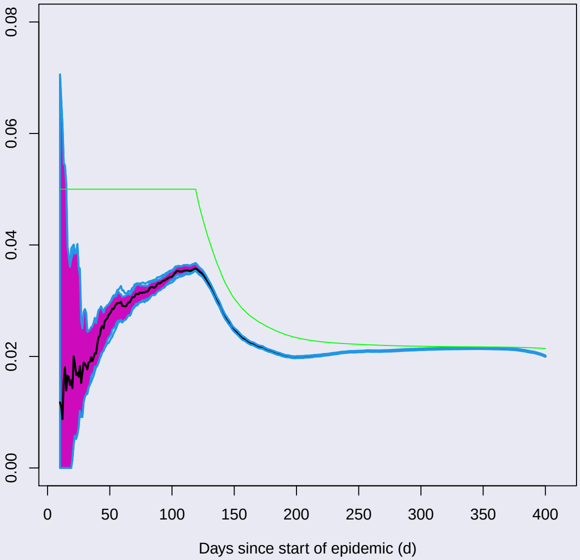
Boxplot funcional del Estimador de Garske.

Letalidad abrupta, c_d de India, sin estimar F.



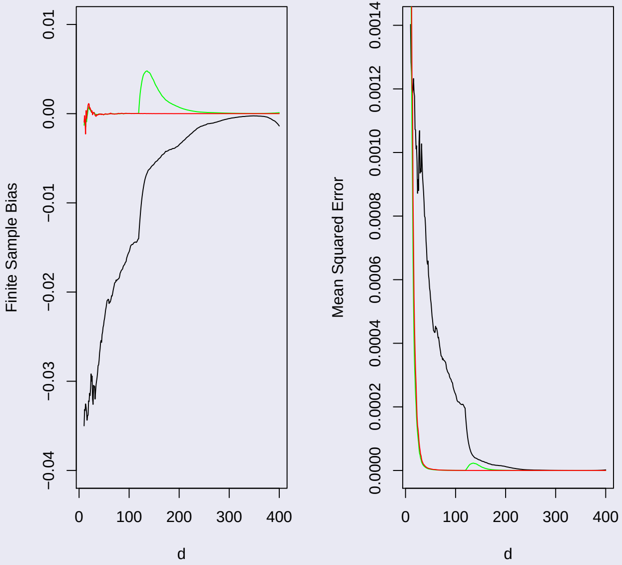
Boxplot funcional del Estimador Naive

Letalidad abrupta, c_d de India, sin estimar F.



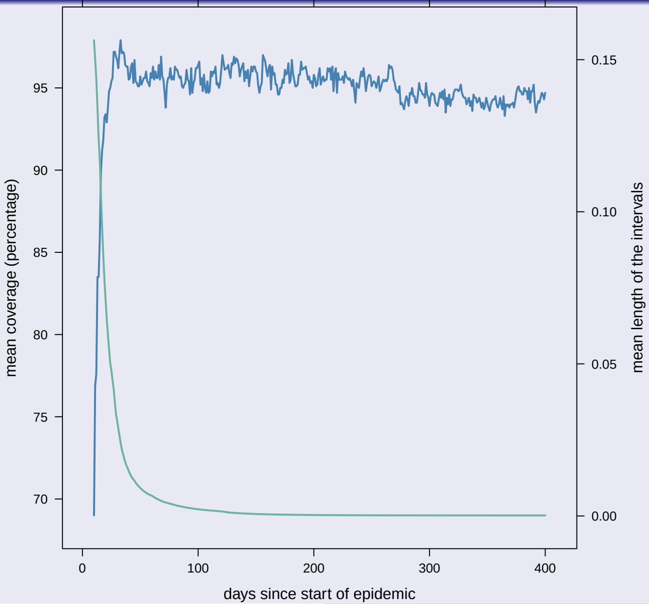
Sesgo y ECM para muestras finitas

Letalidad abrupta, c_d de India, sin estimar F.



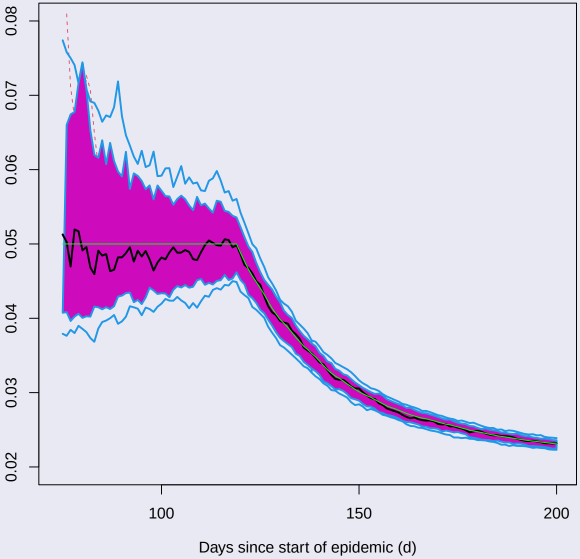
Cubrimiento Empírico de los Intervalos de confianza

Letalidad abrupta, c_d de India, sin estimar F.



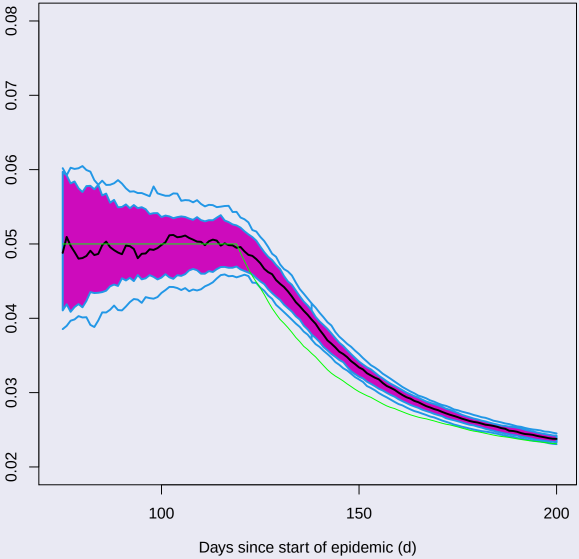
Boxplot funcional del Estimador Nuestro

Letalidad abrupta, c_d de Argentina, estimando F.



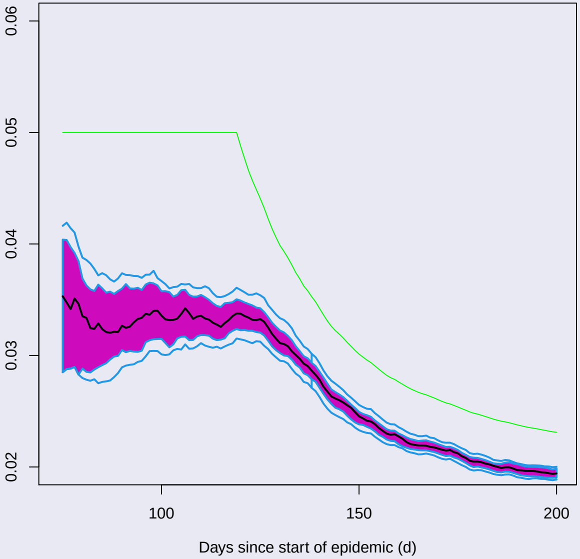
Boxplot funcional del Estimador de Garske.

Letalidad abrupta, c_d de Argentina, estimando F.



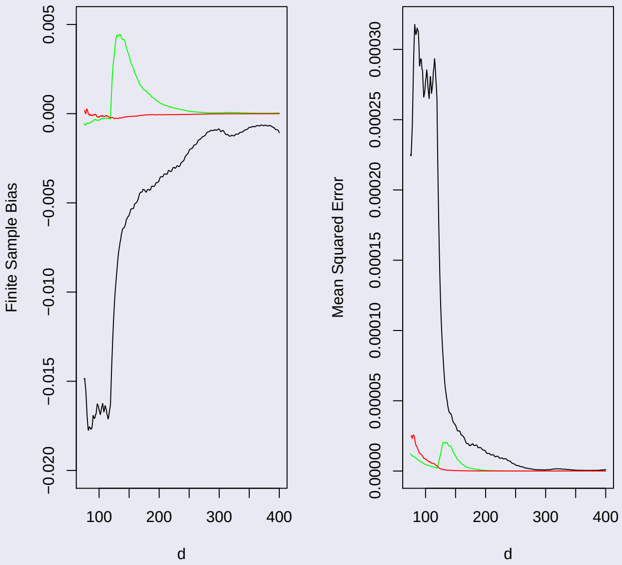
Boxplot funcional del Estimador Naive

Letalidad abrupta, c_d de Argentina, estimando F.



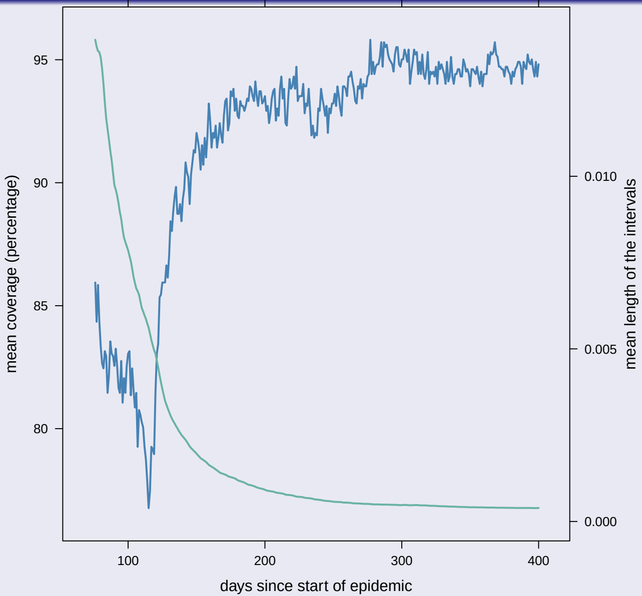
Sesgo y ECM para muestras finitas

Letalidad abrupta, c_d de Argentina, estimando F.



Cubrimiento Empírico de los Intervalos de confianza

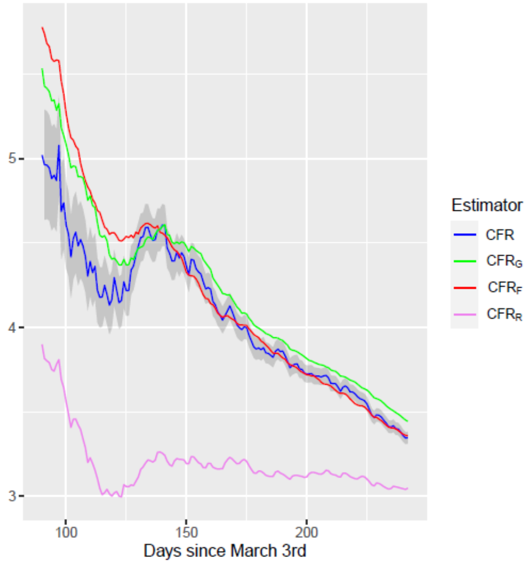
Letalidad abrupta, c_d de Argentina, estimando F.



Aplicación a datos reales en Argentina en 2020

Consideramos los datos de Argentina obtenidos a partir de la página del Ministerio de Salud. Estimamos la Tasa de Letalidad de Covid-19 por diferentes métodos y como F utilizamos la empríca.

Tasa de letalidad de Covid-19 en Argentina desde el 1 de junio hasta el 31 de diciembre de 2020.



Algunas Referencias

1. Billingsley, P., 2008. *Probability and measure.* John Wiley & Sons.

2. Garske, T., Legrand, J., Donnelly, C.A., Ward, H., Cauchemez, S., Fraser, C., Ferguson, N.M., Ghani, A.C., 2009. *Assessing the severity of the novel inuenza a/h1n1 pandemic.* Bmj 339.

Donde ver este trabajo

- Esta presentación está basada en nuestro artículo: *Estimating the case fatality rate of a disease during the course of an epidemic with an application to COVID-19 in Argentina,* disponible en https://arxiv.org/pdf/2109.03087v1.pdf

- Algunos conceptos sobre este tema los desarrollamos para Profesores de nivel Secundario, en el capítulo *Letalidad de Covid-19*, del libro *COVID se enseña*, Ediciones UNGS, Los Polvorines, Bs.As. (Envíado para evaluación)

Agradecimientos

Al grupo de Estadística Aplicada a Covid 19 convocado por el Instituto de Cálculo (FCEyN-UBA) integrado por:

- Ana Bianco
- Daniela Rodriguez
- Diana Kelmansky
- Guillermo Solovey
- Jemina García
- Juan Gonzalez
- Lila Ricci
- Mariela Sued
- Pablo Terlisky
- Pablo Vena

Gracias!!!