

# Análisis Estadístico del COVID-19 en Estados Unidos y Corea del Sur

Isabella Martínez Martínez y Andrés Felipe Florián Quitián.

Universidad del Rosario

Escuela de Ingeniería, Ciencia y Tecnología

*Repositorio:* <https://github.com/felipeflorian/Estadistica/tree/master/Proyecto>

May 19, 2020

## 1 Resumen ejecutivo

A lo largo del 2020 el mundo se ha visto estremecido por una pandemia, fruto de una nueva cepa de coronavirus denominado COVID-19. Así pues, desde un punto de vista epidemiológico, es de particular interés estudiar el crecimiento del mismo mediante modelos estadísticos, así como estudiar los cambios de comportamiento de las poblaciones para luchar en esta coyuntura; comparando a dos potencias mundiales como lo son Corea del Sur y Estados Unidos.

Con un buen manejo de la situación, los casos diarios adoptan una forma de campana de Gauss, mientras que las muertes acumuladas y los tests realizados acumulados se empiezan a estabilizar (debido al fin de la pandemia). Comportamentalmente, la población surcoreana ha adoptado mayor cantidad de hábitos positivos, por otro lado, los estadounidenses se encuentran más preocupados y con una mayor insatisfacción ante las acciones tomadas por su gobierno.

## 2 Introducción y descripción del problema

El control del COVID-19 ha sido determinante para evitar el colapso del sistema de salud, lo que implica una reducción en los casos positivos y en la tasa de mortalidad. Por ello en este proyecto se realizó un análisis acerca de las medidas tomadas y los efectos provocados por estas entre Estados Unidos y Corea del Sur desde enero hasta principios de mayo. Así pues, se le brindó especial importancia a las curvas con respecto a casos diarios, número de test realizados acumulados y número de muertos acumulados; además se realizaron pruebas de hipótesis respecto a cambios comportamentales de la población y su opinión respecto al gobierno y a la crisis. Finalmente,

se estimó puntualmente y mediante intervalos de confianza distintas medias poblacionales de las curvas mencionadas anteriormente.

Esta comparación cobra relevancia con el dato adicional de que tanto Corea del Sur como Estados Unidos tuvieron su primer caso confirmado el 20 de enero de 2020, por lo que ambas naciones han luchado contra el virus durante la misma cantidad de días. Siendo ambos países potencias mundiales pero con culturas y gobiernos diferentes es importante reconocer sus esfuerzos y cómo estos han afectado a la pandemia.

### 3 Datos a usar

Primero se consideraron los siguientes datos, los cuales abarcan aproximadamente los tres primeros meses del transcurso de la pandemia en Estados Unidos y Corea del Sur.

- Total de casos y muertes confirmadas (Diarios y Acumulados).
- Total de pruebas realizadas.
- Cambio de la rutina de las personas.
- Nivel de preocupación.
- Opinión respecto al gobierno.

Además, se tomaron los datos de finales de abril y principios de mayo con el fin de comparar nuestro modelo con los datos reales. Estos datos fueron consultados en *Statista* y *Kaggle*.

## 4 Análisis de datos

### 4.1 Regresiones y modelos propuestos

Dada la estimación de que una persona infectada le puede transmitir el virus a 3 personas más se esperaría que, sin tomar medidas, el comportamiento de los nuevos casos diarios presentara un comportamiento exponencial. Con esto en mente, se revisó el número de casos diarios reportados hasta el 16 de abril en Corea del Sur y Estados Unidos, con el fin de presentar un modelo que se adecuara a los datos. Con los datos hasta este punto se puede concluir que Estados Unidos no ha alcanzado el pico de su contagio todavía.

Para Estados Unidos se concluyó que los datos seguían un comportamiento exponencial y se realizó una regresión lineal en consecuencia, véase Fig.1. La forma general de un modelo ex-

ponencial es  $y = \alpha e^{\beta x}$ , en este caso el modelo es  $y = e^{-2.810111} e^{0.156694x}$ , con  $R^2 = 0.8$  y una correlación de 0.89.

Además, un intervalo de confianza del 95% para los parámetros del modelo es de  $[e^{-3.6577350}, e^{-1.9624866}]$  para  $\alpha$  y  $[0.1401515, 0.1732362]$  para  $\beta$ .

Para los datos de Corea del Sur se realizó una regresión no lineal con el fin de obtener una curva gaussiana. La forma general del modelo utilizado es  $y = a \exp(-\frac{1}{2}(\frac{x-m}{s})^2)$ , y el modelo obtenido es  $y = 636.2375 \exp(-\frac{1}{2}(\frac{x-43.4435}{5.1073})^2)$ , que se puede observar en la Fig.2. Note que se forzó en el modelo que la curva tienda a cero conforme se aleja de su media, ya que el sentido común dice que los datos eventualmente llegarán a cero. Asimismo, se puede ver que el pico de contagio en Corea se dio aproximadamente en el día 40 desde el primer caso.

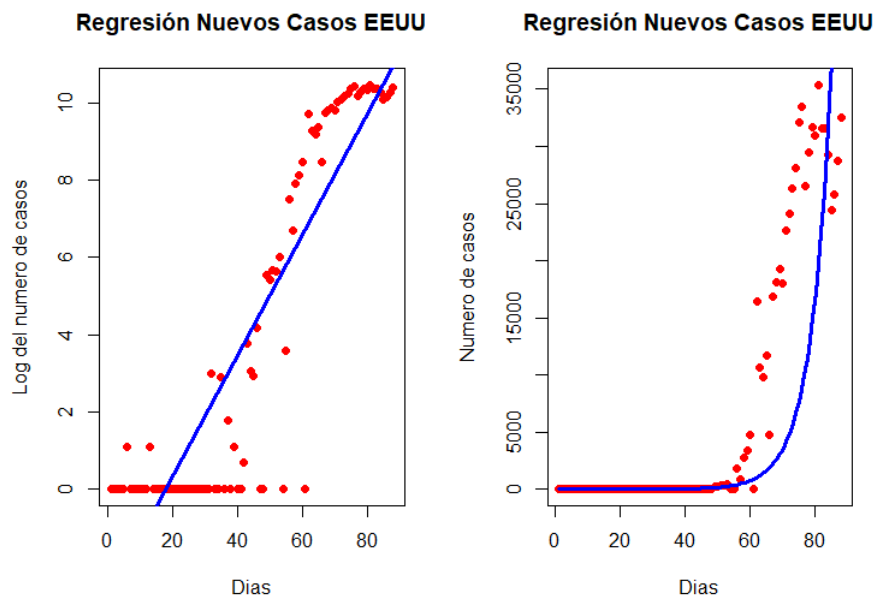


Figure 1: Regresión exponencial para nuevos casos diarios en EEUU

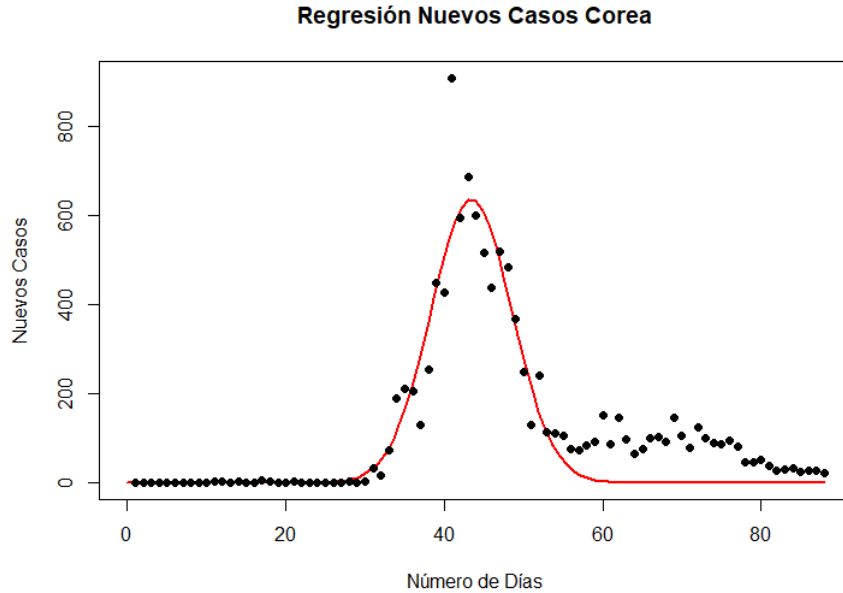
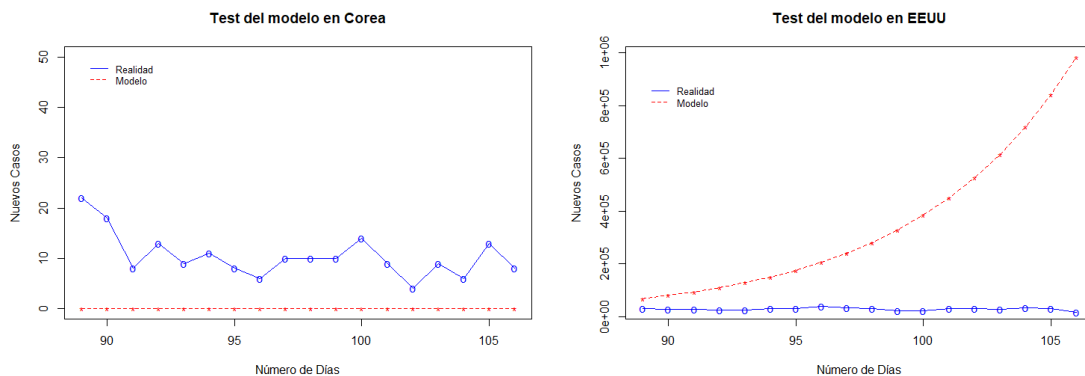


Figure 2: Regresión no lineal para nuevos casos diarios en Corea

Se probaron los modelos establecidos anteriormente con información actualizada al 4 de mayo, es decir, se analizó qué tan acertados son los modelos para los días 89 al 106 desde el primer caso. Note en la Fig.3a que a esas alturas el modelo gaussiano ya es prácticamente cero para nuevos casos, mientras que en la realidad todavía se presentan algunos casos, aunque bajos. El cambio más notorio es el modelo propuesto para Estados Unidos, como se aprecia en la Fig.3b. Observe que el modelo sigue su curva exponencial hasta niveles bastante altos, mientras que la realidad muestra que el gobierno estadounidense está logrando un aplanamiento de la curva poco a poco, con valores de contagios diarios entre 15.000 y 40.000, una mejoría respecto a datos anteriores.



(a) Test del modelo para Corea del Sur

(b) Test del modelo para Estados Unidos

Figure 3: Test de los modelos

Con esta información adicional también se realizaron estimaciones tanto puntuales como intervalares para la media de los nuevos casos en Corea del Sur y Estados Unidos. Se encontró que Corea del Sur presenta en promedio 102 casos diarios mientras que Estados Unidos tiene 11049, una diferencia notoria. Además, los intervalos de confianza de 95% son  $[69.65166, 134.14079]$  para Corea y  $[8452.329, 13645.482]$  para EEUU.

Teniendo en cuenta que la detección temprana del virus, ayuda a tomar las medidas necesarias para detener la propagación del virus. Se realizó una regresión lineal para el número de pruebas diarias realizadas en ambos países.

Para el caso de Estados Unidos se notó que el número de pruebas creció de manera exponencial luego de un mes de que se reportara el primer caso, por ello se realizó un modelo exponencial para el número de pruebas diarias. En donde el modelo obtenido está dado por  $y = e^{5.7666183}e^{0.1459993x}$ . Además,  $R^2 = 0.4673$  y la correlación es 0.6836116.

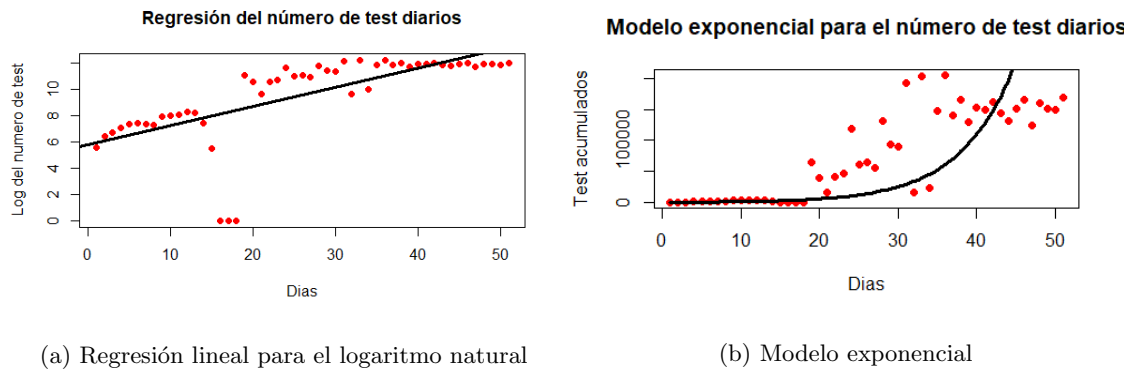


Figure 4: Regresión para el número de pruebas diarias en Estados Unidos.

Las gráficas que se muestran se realizaron con las pruebas realizadas entre el 1 de marzo y el 20 de abril. Por otro lado, se encontró un intervalo de confianza del 95% para los parámetros del modelo los cuales son:  $[e^{4.4296427}, e^{7.1035940}]$  para  $\alpha$  y  $[0.1012508, 0.1907478]$  para  $\beta$ .

De manera analoga se realizó lo mismo para Corea del Sur, en donde su modelo está dado por  $y = e^{5.81226758}e^{0.05512746}$ . Además,  $R^2 = 0.5411$  y su correlación es 0.7355633. Sus gráficas se realizaron con los datos entre el 23 de enero y el 16 de abril.

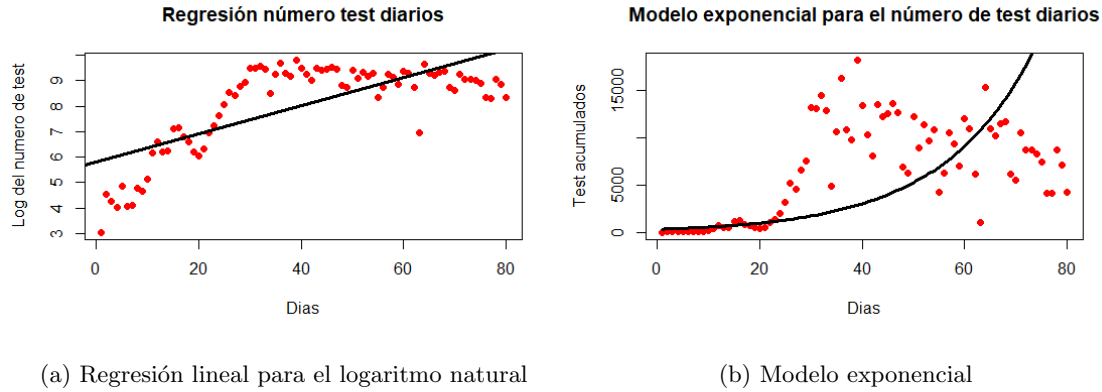


Figure 5: Regresion para el número de tests diarios en Corea del Sur

Por otro lado, se hallaron intervalos de confianza del 95% para sus parámetros los cuales son:  $[e^{5.27868771}, e^{6.34584746}]$  para  $\alpha$  y  $[0.04368236, 0.06657257]$  para  $\beta$ .

Con los datos actualizados hasta el 4 de mayo se probaron los modelos establecidos anteriormente. Se analizó que tan acertados son en los días 52 al 66, 81 al 98, en Estados Unidos y Corea del Sur respectivamente. En la Fig. 6b se evidencia el desfase que tiene el modelo con la realidad. Debido a que, el modelo sigue su curva exponencial, mientras que el número de pruebas diarias en Estados Unidos se mantiene entre los mismos valores. También, en la Fig. 6a se nota que el modelo y la realidad difieren en sus valores. Debido a que este modelo no toma en cuenta el número de casos que se presentan en Corea del Sur, sabemos que estos disminuyeron para este tiempo, por lo cual el número de pruebas realizadas disminuyó con estos.

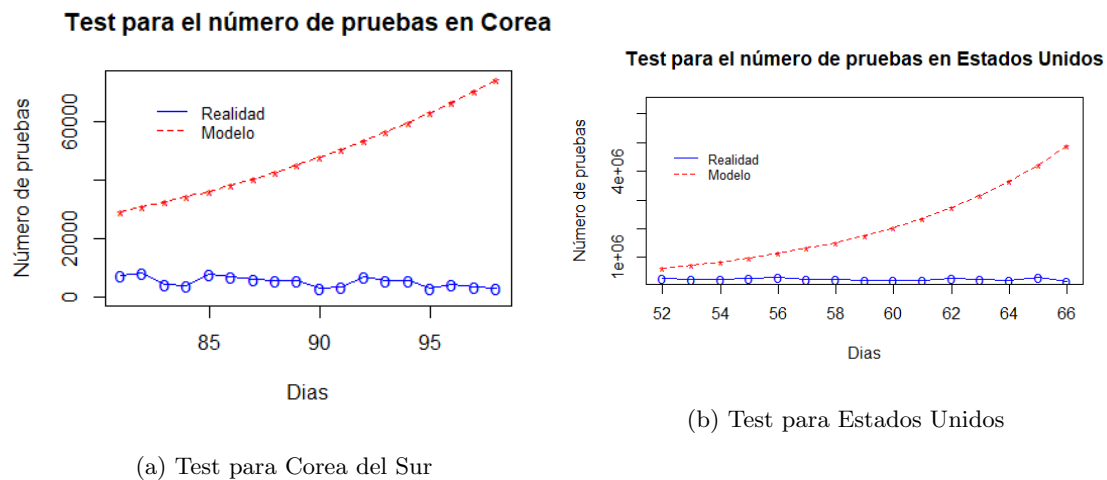


Figure 6: Test de los modelos.

También, se realizaron estimaciones tanto puntuales como intervalares para la media de las

pruebas realizadas en Corea del Sur y Estados Unidos. Para Corea del Sur se encontró que la media de pruebas realizadas es 6486 y un intervalo de confianza del 95% de [5542.149, 7395.014] y en Estados Unidos se encontró que la media de pruebas diarias es 72541 y su intervalo de confianza del 95% es [55248.04, 89835.18].

Además, para ambos países se construyó un modelo para el número de muertes acumuladas y una regresión lineal. Para Corea del Sur se obtuvo que su modelo está dado por  $y = e^{-1.06936857}e^{0.08380922x}$ . Además,  $R^2 = 0.9042$  y su correlación es 0.9265085.

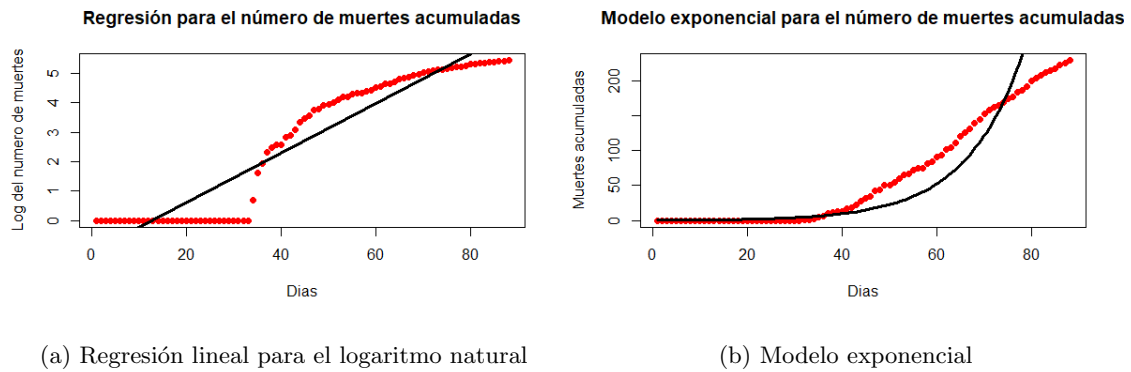


Figure 7: Regresión para el número de muertes acumuladas en Corea del Sur

También, se hallaron intervalos de confianza para sus parámetros  $[e^{-1.36904816}, e^{-0.76968897}]$  para  $\alpha$  y  $[0.07796063, 0.08965781]$  para  $\beta$ .

Asimismo, se obtuvo que para Estados Unidos su modelo es  $y = e^{-3.3646732}e^{-2.202537x}$ . Por otro lado,  $R^2 = 0.8768$  y su correlación es 0.9363871.

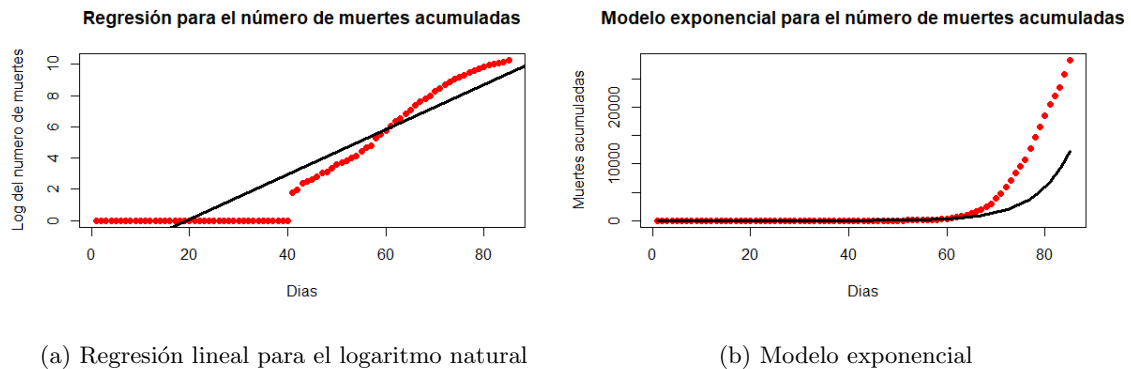
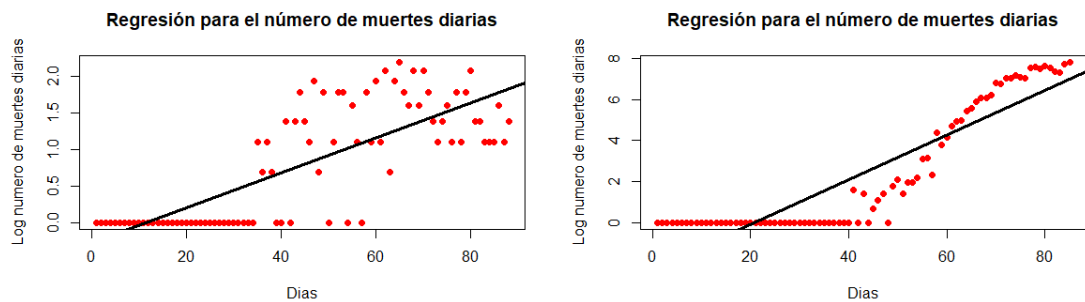


Figure 8: Regresión para el número de muertes acumuladas en Estados Unidos

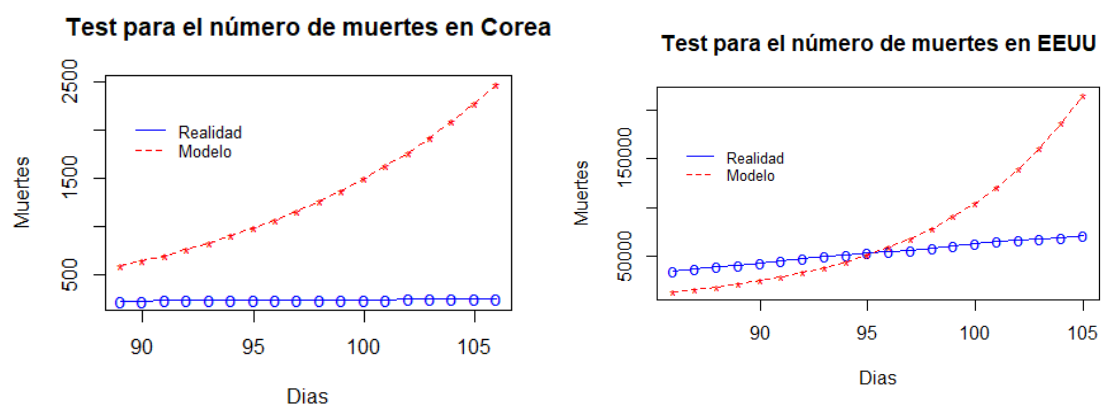
Se encontraron intervalos de confianza para sus parámetros  $[e^{-3.3646732}, e^{-2.202537}]$  para  $\alpha$  y  $[0.1316981, 0.155172]$  para  $\beta$ .

Por otro lado, se realizaron las gráficas respectivas de regresión lineal para el número de muertes diarias en ambos países.



(a) Regresión lineal muertes diarias Corea del Sur (b) Regresión lineal muertes diarias Estados Unidos

Con los datos actualizados se probaron los modelos. Se analizó qué tan acertados son en los días 89 al 106 en Corea del Sur y para Estados Unidos los días 86 al 105. Notamos que en la Fig. 10a el modelo alcanza a considerar 2500 muertes. Sin embargo, el modelo presenta valores muy altos a los reales debido a que el modelo no toma en cuenta la tasa de mortalidad ni los casos críticos que se presentan. Además, notamos que Corea del Sur logró disminuir el número de muertes a tal punto en el que hay días que ya no presentan. Por otro lado, en la Fig. 10b, se nota que el modelo en cierto tiempo esta por debajo de los datos reales. Sin embargo, alrededor del día 90 el modelo se disparó y continuo su curva exponencial. Esto nos dice que así las muertes en Estados Unidos hayan aumentado a tal punto que ya presentan mas de 50000 muertes, han logrado disminuir el número de muertes diarias ya que el aumento de las muertes acumuladas ha bajado después del día 90.



(a) Test muertes acumuladas Corea

(b) Test muertes acumuladas EEUU



## 4.2 Pruebas de Hipótesis

Mediante encuestas realizadas a ambas poblaciones podemos analizar la diferencia en cuanto a comportamiento y opinión que presentan. Luego, es prudente hallar el valor  $p$  para comprobar la veracidad de las proporciones presentadas y realizar pruebas de hipótesis comparando las proporciones de ambas poblaciones.

En cuanto al nivel de preocupación presentado ante el COVID-19 se cuenta con una encuesta realizada entre el 3 y 5 de marzo para los surcoreanos y el 13 y 16 de marzo para los estadounidenses. En el caso de Corea del Sur, el tamaño muestral fue de 1000 personas, con el 74% diciendo estar preocupados ante la posibilidad de infectarse, mientras que en Estados Unidos se le preguntó a 1968 personas y el 80% dijo estar preocupados. Así pues, con una hipótesis nula de que las proporciones presentadas son iguales al 50% con una alternativa de que son mayores para ambas poblaciones, se obtiene que con una certeza de prácticamente el 100% se puede afirmar que la mayoría de surcoreanos y estadounidenses están preocupados respecto al COVID-19.

Además se realizó una prueba de hipótesis comparando estas proporciones. Nuevamente, con una certeza casi del 100% se afirma que los estadounidenses están más preocupados que los surcoreanos respecto a la pandemia global.

En este orden de ideas cabe preguntarse sobre la opinión de estas poblaciones respecto a las acciones tomadas por sus respectivos gobiernos para luchar contra el virus. En relación con esto se cuenta con encuestas realizadas del 28 de febrero al 2 de marzo para los surcoreanos, así como una encuesta diaria del 23 de marzo al 15 de abril para los estadounidenses. Con el fin de escoger solo un valor para esta última encuesta y utilizando el hecho de que las proporciones de satisfacción no varían mucho, se escogió un valor promedio.

Luego, al encuestar a 1000 surcoreanos se concluyó que el 42% están satisfechos con el gobierno de Moon Jae-in, mientras que la muestra de 7700 estadounidenses dio como resultado que el 39% de los estadounidenses tienen una opinión positiva de las acciones tomadas por Donald Trump. También aquí se puede afirmar al 100% que la mayoría de surcoreanos y estadounidenses desapruaban las acciones tomadas por sus respectivos gobiernos. Asimismo, se realizó una prueba de hipótesis para comparar ambas proporciones, y se concluye que se puede afirmar con un 97% de confianza que los estadounidenses están más insatisfechos con su gobierno que los surcoreanos.

Los cambios de comportamiento han sido determinantes para detener la propagación del virus, por ello se realizaron dos pruebas de hipótesis respecto a cambios de hábitos en las poblaciones surcoreana y estadounidense. Además, se halló el valor  $p$  el cual es pertinente para dar una conclusión. Para esto se tienen en cuenta encuestas para los dos países. Para Corea del Sur se tomó una encuesta realizada entre el 28 de febrero y 3 de marzo a 1000 personas y para Estados Unidos una encuesta realizada los días 10 y 11 de marzo a 1635 personas.

Se tomó en cuenta si estas poblaciones dejaron de concurrir a lugares en los que se aglomera una gran cantidad de gente. El 96% de la población surcoreana dejó de asistir a estos lugares. Sin embargo, solo el 37% de la población estadounidense evade asistir a estos lugares. Así pues con una hipótesis nula de que los datos presentados son iguales al 0.5 para ambos casos y una hipótesis alternativa de que estos datos son mayores para la población surcoreana y menores para la estadounidense. Se puede afirmar que casi el 100% de la población surcoreana evade los lugares llenos de gente. Por otro lado, se puede afirmar que menos del 50% de la población estadounidense cumple este hábito.

Asimismo, se realizó una prueba de hipótesis comparando estos datos, en donde podemos concluir que los surcoreanos se adaptaron a este cambio de hábito. Sin embargo, los estadounidenses no se adaptaron al mismo.

Además, tomamos en cuenta si las poblaciones empezaron a comprar más elementos de supermercado los cuales se pueden almacenar. La encuesta muestra que el 35% y el 15% de las poblaciones surcoreana y estadounidense adoptaron esta medida respectivamente. Así con una hipótesis nula de que los datos son iguales al 0.5 y una alternativa de que estos son menores, podemos afirmar que tanto la población surcoreana como la estadounidense no ha adoptado este hábito. Por último, realizando una prueba de hipótesis comparando estas proporciones y se concluye que un mayor porcentaje de la población surcoreana adoptó este hábito.

## 5 Conclusiones

1. Eventualmente las curvas de los casos diarios, tanto para Corea del Sur como para Estados Unidos adoptarán una forma de campana de Gauss. Para Corea es más evidente pues alcanzó esa forma en menor cantidad de tiempo. Estados Unidos, a la fecha, está empezando a aplanar la curva y por ende está apenas empezando a parecerse a una campana.
2. La media de casos diarios es para Corea del Sur de 102, mientras que para Estados Unidos es de 11049. Esto es el 0.0001% de la población surcoreana y el 0.003% de la población estadounidense, por lo que Estados Unidos tiene más contagios diarios que Corea del Sur, en proporción.
3. Tanto para Corea del Sur como para Estados Unidos se observa que la mayoría de sus habitantes está en desacuerdo con las acciones tomadas por su gobierno ante el COVID-19. Los estadounidenses están más insatisfechos con Trump que los surcoreanos con Moon.
4. De manera similar, la mayoría de los habitantes de ambas poblaciones se encuentran preocupados por un posible contagio del virus. Los estadounidenses están más preocupados que los surcoreanos.

5. La población surcoreana adoptó mayores cambios de hábitos como el evadir ir a lugares donde se encuentra una gran aglomeración de gente. Además, una pequeña parte de la población empezó a comprar más artículos de supermercado. Por otro lado, los estadounidenses no presentaron grandes cambios en sus hábitos, como el evitar ir a lugares concurridos los cuales llegan a ser un foco de la infección.
6. La media de pruebas diarias para Corea es de 6486 teniendo en cuenta que para la llegada del pico de la infección el número de pruebas superó las 10000 pruebas realizadas y este número decreció a medida que la curva bajaba. Para Estados Unidos la media de pruebas diarias es de 72541.
7. Se observó que el número de muertes acumuladas en Corea del Sur comienza a ser estable, esto debido a que hay días en los que ya no presentan muertes. Por otro lado, en Estados Unidos el número de muertes acumuladas ha dejado de crecer de manera alarmante. Sin embargo, aun presentan miles de muertes diarias.

## References

- [1] Wackerly, Mendenhall, Scheaffer. *Mathematical Statistics with Applications, 7th Ed.* Thomson Brooks/Cole, 2008.
- [2] Statista
- [3] Kaggle