

UNIVERSIDAD NACIONAL AUTÓNOMA DE NICARAGUA
UNAN-LEÓN
FACULTAD DE CIENCIAS QUÍMICAS
CARRERA DE FARMACIA



Monografía para Optar al Título de Lic. Químico Farmacéutico

Aplicación de la ley de Benford para verificar la validez de los casos de COVID-19 reportados por la Universidad Johns Hopkins en el periodo comprendido de enero - junio 2020.

Autores:

Br. Julissa Fernanda Matamoros López.

Br. Salvador Enrique Pastora Altamirano.

Tutor:

Ing. Roberto de Jesús Torrez Barrera.

León, julio 2022

“A la Libertad por la Universidad”

UNIVERSIDAD NACIONAL AUTÓNOMA DE NICARAGUA
UNAN-LEÓN
FACULTAD DE CIENCIAS QUÍMICAS
CARRERA DE FARMACIA



Aplicación de la ley de Benford para verificar la validez de los casos de COVID-19 reportados por la Universidad Johns Hopkins en el periodo comprendido de enero - junio 2020.

Autores:

Br. Julissa Fernanda Matamoros López.

Br. Salvador Enrique Pastora Altamirano.

Tutor:

Ing. Roberto de Jesús Torrez Barrera.

Firmas

León, julio 2022

“A la Libertad por la Universidad”



DEDICATORIA

A DIOS: Por su infinita misericordia cada día a lo largo de mi vida y por ayudarme a cumplir uno de mis mayores sueños, culminando mi carrera Universitaria

A MIS PADRES: Julio Matamoros e Ivania López, por enseñarme a salir adelante, por hacer las cosas correctas, por su apoyo incondicional hasta el final, valores, esfuerzos y amor infinito a lo largo de mi vida, por apoyarme en cada uno de mis triunfos y logros.

En memoria de: ING Julio Matamoros Montenegro, A las tias: Coco y Cristina Montenegro, a mis tíos: Marvin y Freddy Matamoros y a mi querido sobrino: Santiago Nicolás Matamoros. Todos ellos (Q.D.E.P) de esta manera honrado su memoria para poder terminar uno de mis más grandes sueños y se que de ellos también lo era, por depositar mucho empeño y confianza hacia a mí.

A todos ellos mil gracias.

BR: Julissa Fernanda Matamoros López.



DEDICATORIA

El presente trabajo investigativo lo dedicamos principalmente a Dios, por ser el inspirador y darnos fuerza para continuar en este proceso de obtener uno de los anhelos más deseados.

A nuestros padres, tíos y abuelos por su amor, trabajo y sacrificio en todos estos años, gracias a ustedes hemos logrado llegar hasta aquí y convertirnos en lo que somos. Ha sido el orgullo y el privilegio de ser sus hijos, son los mejores padres.

Nuestras hermanas (os) por estar siempre presentes, acompañándonos y por el apoyo moral, que nos brindaron a lo largo de esta etapa de nuestras vidas.

A todas las personas que nos han apoyado y han hecho que el trabajo se realice con éxito en especial a aquellos que nos abrieron las puertas y compartieron sus conocimientos.

A la Br: Julissa Matamoros por ser mi mayor motor durante estos años de carrera y ser la mejor compañera, novia y equipo para realizar cada logro de esta etapa universitaria y de preparación al profesionalismo.

BR. Salvador Enrique Pastora Altamirano.



AGRADECIMIENTO

A DIOS: por ser quien nos da vida y fuerzas para salir adelante, tanto como sabiduría e inteligencia para poder empaparnos del saber y culminar nuestra carrera universitaria, como profesionales de salud dispuestos al servicio de nuestra comunidad y país.

A NUESTROS PADRES: por su arduo trabajo, por su empeño para que seamos personas de bien, colmándonos de valores y por su apoyo incondicional, porque sin ellos no pudiésemos lograrlo y culminar nuestros estudios universitarios.

A NUESTRO TUTOR: MSC Roberto Torrez, por su apoyo y sobre todo su disposición para compartir y transmitir sus conocimientos brindados hacia nosotros, ya que por su apoyo y asesoría pudimos culminar nuestra monografía.

A NUESTROS MAESTROS: licenciados de la carrera de Farmacia, por impartirnos conocimientos y enseñanzas de la mano a la par de valores.

A Nuestra querida Alma Mater UNAN-LEON. Por dicha formación profesional, educándonos con principios y altos grados de conocimientos y competitividad científica y calidad humana.

A mi compañero: Salvador Pastora, por todos estos años de convivencia de carrera y por la realización y experiencia de este trabajo monográfico.

Br: Julissa Fernanda Matamoros López



AGRADECIMIENTO

Agradecemos a Dios por bendecirnos la vida, por guiarnos a lo largo de nuestra existencia, ser el apoyo y fortaleza en aquellos momentos de dificultad y de debilidad.

Gracias a mis padres: DR Rafael Pastora y Lic. Judith Altamirano, por ser los principales promotores de nuestros sueños por confiar y creer en nuestras expectativas, por los consejos, valores y principios que nos han inculcado.

Agradecemos a nuestros docentes de la Universidad Nacional Autónoma de Nicaragua León por haber compartido sus conocimientos a lo largo de la preparación de nuestra profesión, de manera especial al MSC: Roberto Torrez, tutor y maestro de nuestro proyecto de investigación quien nos ha guiado con su paciencia, y su rectitud como docente, y a las personas que nos apoyaron durante nuestro trabajo.

Br. Salvador Enrique Pastora Altamirano.



RESUMEN

Ante la emergencia sanitaria del COVID-19, se ha hecho más evidente la importancia de tener datos confiables, la no veracidad de la información puede ocasionar problemas de inconformidad por parte de la población. En este estudio utilizamos la ley de Benford para detectar inconsistencia de casos reportados de COVID-19 en una muestra aleatoria en el periodo comprendido de enero – junio del 2020.

La Ley de Benford, es una herramienta estadística que, logra demostrar que los números tienen un comportamiento regular cuando estos se generan de manera natural.

La muestra se obtuvo de una población de 210 países, que reportaron casos de COVID-19 en la plataforma web de la Universidad de Johns Hopkins, el tipo de muestreo aplicado fue muestreo aleatorio simple, para el cálculo del tamaño de la muestra, aplicamos la fórmula de poblaciones finita, para selección de la muestra y tratamiento de los resultados se hizo uso de la hoja de cálculo de Microsoft Excel 365.

Aplicamos pruebas de inferencia estadística; Z (para comprobar la desviación de cada dígito con respecto a la ley de Benford), Chi-cuadrada (para comprobar la Bondad de ajuste de los datos reportado) y la prueba arco mantisa (como segunda prueba de confirmación del ajuste de los datos).

El Estadístico $Z_{\alpha=0.025}$ demostró que 4 países no se desvían de la ley en ninguno de sus dígitos y 10 se desviaban en por lo menos en un dígito.

Las pruebas Chi-cuadrado y arco mantisa indican que 8 países se ajustan perfectamente a la ley de Benford.

Finalmente, aunque algunos países no cumplen la Ley de Benford según los estadísticos de prueba aplicados no necesariamente indican que existe inexactitud



en el conjunto de datos analizados, simplemente nos dan alertas de donde poner mayor atención en la verificación de estos.



CARTA DE AUTORIZACIÓN DEL TUTOR

A quién pueda interesar

Por la presente hago constar que he asesorado, supervisado, emitido opiniones y he leído el proyecto de investigación titulado Aplicación de la ley de Benford para verificar la validez de los casos de COVID-19 reportados por la Universidad Johns Hopkins en el periodo comprendido de enero - junio 2020, presentado por los Bachilleres: Julissa Fernanda Matamoros López, carnet 16-15130-0 y Salvador Enrique Pastora Altamirano, carnet: 16-02181-0 para aspirar al grado de Licenciados Químicos Farmacéuticos, otorgado por la Facultad de Ciencias Químicas, Carrera de Farmacia, de la Universidad Nacional Autónoma de Nicaragua, UNAN-León, por lo anterior expuesto y en mi condición de tutor académico hago constar que los objetivos del proyecto de investigación fueron desarrollado y alcanzado satisfactoriamente y considero finalizado el proyecto de investigación, y autorizo su presentación a fin de que pueda ser juzgado y calificado por el tribunal correspondiente.

Sin otro particular a que hacer referencia, me despido de usted.

Atentamente;

Ing. Roberto Jesus Torrez Barrera
Tutor Académico



ÍNDICE

INTRODUCCIÓN	1
OBJETIVOS.....	3
General:.....	3
Específicos:.....	3
MARCO TEÓRICO	4
Comportamiento epidemiológico de COVID19 durante la fase inicial de la Pandemia	4
Ley de Benford.....	5
Datos que satisfacen la Ley de Benford.....	6
Datos que no satisfacen la Ley de Benford.....	6
Pruebas de bondad de ajuste	7
Estadístico Z	7
DISEÑO METODOLÓGICO.....	13
Tipo de estudio	13
Área de estudio	13
Población de estudio.....	13
Tamaño de Muestra	13
Tipo de Muestreo:	14
Procedimiento de Muestreo	14
Criterios de inclusión:.....	14
Criterios de exclusión:.....	14
Fuente de información	14
Consideraciones éticas:	14
Método de recolección de datos	15
Plan de análisis	15
Variables:	16
Cruce de variables:	16
OPERACIONALIZACIÓN DE LAS VARIABLES	17
RESULTADOS Y DISCUSIÓN	18
CONCLUSIONES	30



RECOMENDACIONES.....	31
BIBLIOGRAFÍA	32



INTRODUCCIÓN

“Los números suelen comenzar más frecuentemente por «1» que por cualquier otro dígito”

Quien primero se dio cuenta, en 1881, de este fenómeno fue el matemático y astrónomo Simón Newcomb¹, al percatarse de que las páginas de los primeros dígitos, en las tablas de logaritmos, estaban más desgastadas que las páginas de los últimos dígitos, dicha observación lo llevo, mediante un breve e ingenioso razonamiento, aunque sin presentar realmente un argumento formal ni fórmula matemática, a enunciar verbalmente la siguiente ley logarítmica: “La probabilidad de ocurrencia de números es tal que las mantisas de sus logaritmos son equiprobables”.

En 1938, el físico Frank Benford², se dio cuenta del mismo patrón. Entusiasmado por el descubrimiento estudió datos extraídos del mundo real (constantes, magnitudes físicas, longitud de ríos, estadísticas de béisbol, direcciones de personas, incluso cifras sacadas de portadas de revistas). Comprobó que la probabilidad de que un número en una serie de datos comience por el dígito d es $P[d] = \log_{10}(1 + 1/d)$, llevándolo a postular la llamada “Ley de los números anómalo de Benford”, que establece que los datos numéricos extraídos del mundo real, la primera cifra que se espera que aparezca con más frecuencia que el resto de los números es el 1. El análisis de Benford era una prueba de la existencia de la ley, pero tampoco fue capaz de explicar bien por qué era así.

El matemático, Roger Pinkham³ (1961), dio el primer paso para explicar esta curiosa relación, siguiendo el siguiente razonamiento “la distribución de las frecuencias de dígitos debía ser invariante frente a cambios de escala”, es decir, la frecuencia de dígito es invariante tanto si la calculamos en unidades de euros, dólares, dinares o dracmas, o si medimos en pulgadas o metros etc. Aportando pruebas que iban confirmando que la ley de Benford realmente existe.



Una aplicación importante de la ley de Benford se encuentra en la detección de fraudes fiscales (Mark Nigrini 1994)^{4,5}, por lo que varios analistas han sugerido que esta prueba forense podría utilizarse para detectar la posibilidad de manipulación en otros tipos de datos socioeconómicos.

En años recientes algunos autores, tales como: Alexandra Hurtado⁶ (2020) y Raúl Isea⁷ (2020), han empleado esta metodología para evaluar la calidad de sus datos, el primero utilizó la ley como una herramienta que le permitiera comprobar el desempeño nacional de la calidad de los datos durante la vigilancia de la salud pública de casos acumulados de COVID-19 en el primer semestre en Colombia, el segundo la aplicó con el objetivo de verificar la calidad de los datos aportados por cada país durante la determinación de un modelo matemático capaz de reproducir la dinámica de transmisión del nuevo coronavirus en grupos humanos.

Como podemos observar la ley ha sido utilizada por los diferentes investigadores para validar la hipótesis relacionada con la confiabilidad de los datos, la confirmación de esta hipótesis de confiabilidad es crucial para los procesos de toma de decisiones tanto operativa como estratégica, en tiempo de pandemia, contar con datos confiables les permite a las organizaciones de salud de cada país orientar los esfuerzos humanos y económico para comprender y controlar más eficientemente la enfermedad.

En este estudio aplicamos la ley de Benford y estadísticos de prueba para determinar el número de países que están reportando, de manera confiables sus cifras de casos de COVID-19 en el periodo comprendido de enero - junio 2020. Como profesionales de la salud, es de gran importancia ya que día a día aparecen nuevos datos de caso de COVID-19 cuya credibilidad debe estar acompañada por datos estadísticos confiables y al mismo tiempo saber que existe una Ley estadística, muchas veces ignorada por la comunidad científica, que permite evaluar la veracidad de la información estadística suministrada.



OBJETIVOS

General:

- Aplicar la ley de Benford para verificar la validez de los casos de COVID-19 reportados por la Universidad Johns Hopkins en el periodo comprendido de enero - junio 2020.

Específicos:

- Verificar que los números de casos de COVID-19 reportados por la Universidad de Johns Hopkins sigue una distribución según la ley de Benford,
- Aplicar pruebas de inferencia estadística para evaluar la conformidad de la ley de Benford a los datos reportados de COVID-19.
- Indicar el número de países que están reportando cifras confiables de casos positivos de COVID-19



MARCO TEÓRICO

Comportamiento epidemiológico de COVID19 durante la fase inicial de la Pandemia

La existencia de los virus se sospechó desde el momento del despegue de la microbiología, a partir de las investigaciones de los grandes pioneros de esta especialidad en la segunda mitad del siglo XIX. En la actualidad, este término se ha convertido en un “conocido indeseable”, pues hace meses que una terrible pandemia, COVID-19, ocasionada por el virus 2 del síndrome respiratorio agudo grave (SARS-Cov-2) está causando pérdidas humanas.

Todo comenzó en la lejana provincia China de Hubei específicamente en su capital, la ciudad de Wuhan, en el mes de diciembre de 2019, cuando se reportaron casos de personas enfermas con un tipo de neumonía desconocida. En los primeros días de enero de 2020 la Organización mundial de la salud (OMS) informó sobre la existencia de un conglomerado de casos de neumonía sin fallecimientos, en la ciudad de Wuhan, el virus se trasladó fuera de las fronteras del gigante asiático. El 30 de enero de 2020, el director general de la OMS, convocó al comité de Emergencias que llegó a un consenso: declarar a la infección por SARS-Cov-2 una emergencia de salud pública de interés internacional. La OMS informó, el propio 30 de enero de 2020, la existencia de un total de 7818 casos confirmados en todo el mundo, la mayoría de ellos en China. El 11 de febrero el nombre de la enfermedad cambió oficialmente a COVID-19 y un mes más tarde, el 11 de marzo de 2020, profundamente preocupadamente de la enfermedad, por su gravedad y por los niveles también impresionantes de inacción; la OMS determinó en su evaluación que COVID-19 podía caracterizarse como pandemia.

Hasta el día 29 de mayo se reportaron en el mundo 5 millones 776 mil 934 casos confirmados para una letalidad 6,23% con presencia en 185 países. En América, el primer caso positivo del nuevo coronavirus se corroboró el 21 de enero. Con prontitud, el virus abarcó todo el continente. Debido al reciente surgimiento de esta



enfermedad, la situación de alarma mundial encontramos el incremento constante de casos que se hace que la prevención y el control de la COVID-19 sean complejos es necesario conocer el comportamiento epidemiológico para el desarrollo de acciones de salud dirigidas a la prevención, control, diagnóstico y tratamiento oportuno y efectivo de la enfermedad, así como reducir la mortalidad de esta misma.

Ley de Benford

Probabilidad de ocurrencia de números es tal que las mantisas de sus logaritmos son equiprobables

La ley de probabilidad de ocurrencia de números es tal que las mantisas de sus logaritmos son equiprobables. (Martínez Gámez & Canizales Rivera, 2009)⁷.

1. Probabilidad (primer dígito significativo = d_1)

$$p[d_1] = \log_{10} \left(1 + \frac{1}{d_1} \right), \text{ donde } d = 1, \dots, 9 \text{ Ec. 1}$$

2. Probabilidad (segundo dígito = d_2)

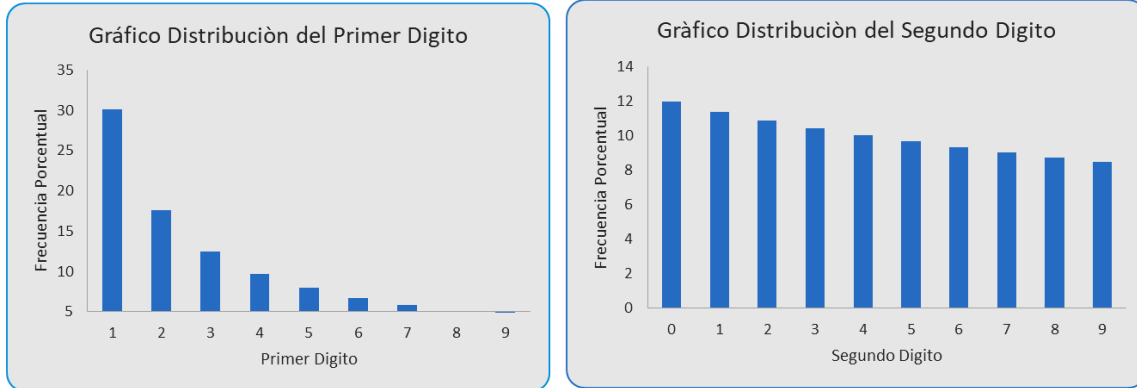
$$p[d_2] = \sum_{d_1=1}^9 \log_{10} \left(1 + \frac{1}{(10d_1 + d_2)} \right), d_1 = 0,1, \dots, 9 \text{ Ec. 2}$$

Tabla 1. Frecuencia de aparición de los primeros dos dígitos significativos dado por Simón Newcomb

d	0	1	2	3	4	5	6	7	8	9
P($d_1=d$)		0.3010	0.1761	0.1249	0.0969	0.0792	0.0669	0.0580	0.0512	0.0458
P($d_2=d$)	0.1197	0.1139	0.1088	0.1043	0.1003	0.0967	0.0934	0.0904	0.0876	0.0850



Gráfico 1. Frecuencia de aparición de los primeros dos dígitos significativos por Newcomb



Datos que satisfacen la Ley de Benford. (Martínez Gámez & Canizales Rivera, 2009)⁷. A continuación, se enuncian una serie de variables cuyos datos se han probado y cumplen con la Ley de Benford, entre estos se tienen:

- Estadísticas de béisbol.
- Constantes y magnitudes físicas y químicas.
- Poblaciones.
- Pagos de impuesto sobre la renta.
- En la desintegración de las partículas radioactivas alfa, se ha encontrado un comportamiento logarítmico que satisface la ley de Benford.
- Magnitudes económicas, sociales, entre otras.

Datos que no satisfacen la Ley de Benford. (Martínez Gámez & Canizales Rivera, 2009)⁷.

Los datos que no siguen la Ley de Benford, son los siguientes:

- Datos provenientes de distribuciones uniformes (loterías)
- Datos sobre edades de las personas (distribuciones normales)
- Números telefónicos, datos de identidad



- Números procedentes de evaluar funciones cuadráticas, raíces, entre otras.

Pruebas de bondad de ajuste de una distribución observada a la ley de Benford. (Mark J. Nigrini, 2012)⁵

La ley de Benford especifica que en una colección de números de dígitos no deben ocurrir con igual frecuencia, si no que el 1 aparece como dígito inicial más frecuente, seguido del 2, etc. hasta el 9 que es el menos probable, un comportamiento muy similar se presenta para el segundo y demás dígitos sólo que en esta vez la diferencia entre las probabilidades de dicho dígito es menos apreciables.

Sin embargo, podríamos plantearnos la siguiente pregunta ¿Cómo podríamos detectar o saber si una colección de números se ajusta o no a la Ley de Benford?, en vista que la ley no es más que una distribución de probabilidad la respuesta que se esperaría en muchas personas es simplemente graficar la frecuencia de apariciones de los dígitos y compararla con la Ley de Benford y luego observar que tan alejados o cercanos estén ambas distribuciones,

sin embargo, realizar tal actividad podría no ser lo conveniente debido a que personas con diferentes puntos de vista podría llegar a diferentes conclusiones; por otra parte existen estadísticos que nos brindan una serie de pruebas de bondad de ajuste capaces de indicarnos si una colección de números se ajusta o no a la Ley de Benford; de hecho aunque no se especifique formalmente que la detección de datos “inventados” es una prueba de hipótesis (en la cual la hipótesis nula es que no existe manipulación), en base a ello se decide si ha existido o no “manipulación”.

Estas pruebas simplemente se encargan de examinar que también se ajusta (alejan) los datos observados a la distribución teórica de Benford, el Dr. Mark Nigrini, el cual es uno de los expertos y pioneros en el tema propone las siguientes pruebas: Z, χ^2 (Chi-cuadrada) y arco mantisa, entre otras.

Estadístico Z (Mark J. Nigrini, 2012)⁵

Esta prueba de bondad de ajuste recoge las desviaciones de los datos observados versus la Ley de Benford para cada uno de los dígitos. De este modo, permite



comprobar si la frecuencia de aparición de cualquier dígito (0, 1, 2,...9) en una posición dada (primero, segundo, etc.), es mayor de la esperada.

El estadístico Z proporcionado por Nigrini está definido por:

$$Z_i = \frac{|P_0^i - P_e^i| - \frac{1}{2n}}{\sqrt{P_e^i \left(\frac{1 - P_e^i}{n} \right)}} \quad \text{Ec. 3}$$

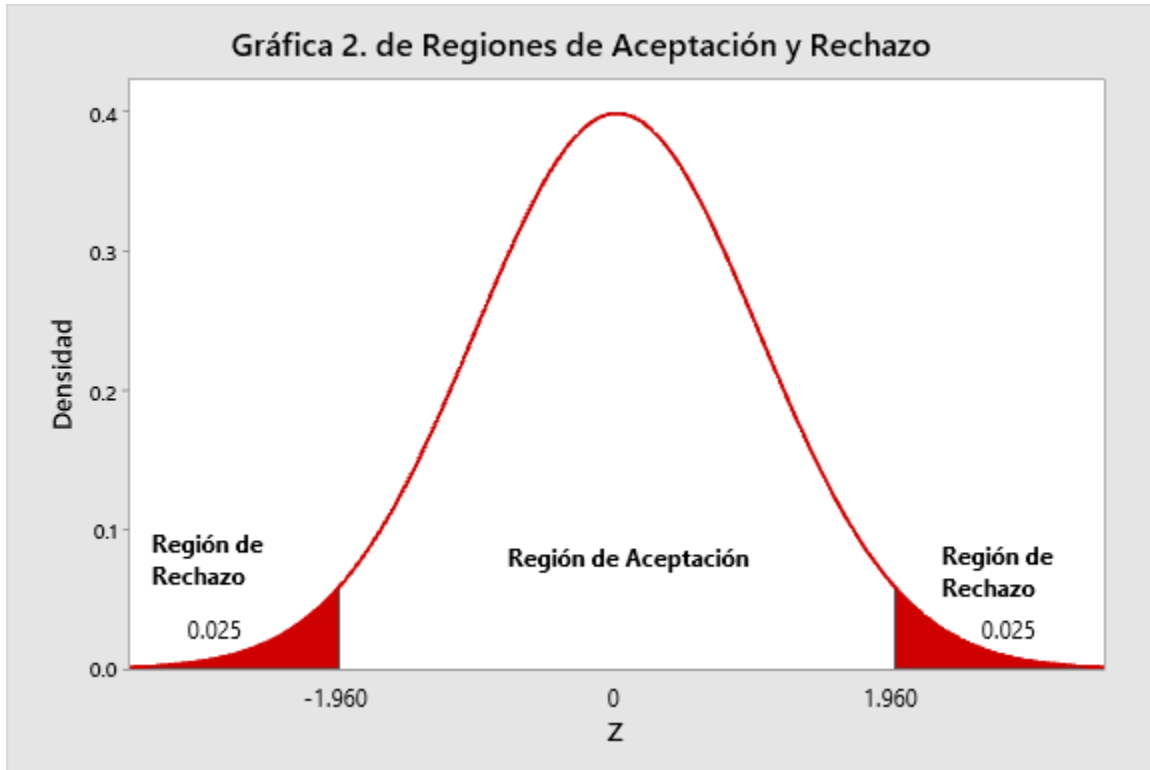
Donde P_e^i denota la proporción esperada para el i-ésimo valor del dígito en estudio.

P_0^i denota la proporción observada para el i-ésimo valor del dígito en estudio.

n es el número de observaciones.

El término $\frac{1}{2n}$ es una corrección de continuidad que únicamente se utiliza cuando el valor es menor que el primer término del numerador. Las proporciones esperadas y observadas deben obtenerse a partir de la correspondiente función de distribución obedeciendo el tipo de prueba que se realiza (primer dígito, segundo, etc.).

El gráfico 2. Proporciona las regiones críticas que debemos considerar para estudiar si la proporción observada es igual a la esperada. Si el valor de Z cae dentro del intervalo de -1.96 a 1.96 podemos concluir que el dígito se ajusta perfectamente a una distribución de la ley de Benford. En el caso que este valor sea menor a -1.96 o mayor a 1.96, el estadístico de prueba nos proporciona una certeza del 95% de confianza que estos dígitos no se ajusta a la distribución de la ley.



Estadístico X^2 (Chi-Cuadrada) (Mark J. Nigrini, 2012)⁵

La prueba X^2 , calcula la bondad de ajuste de todos los dígitos respecto a lo esperados por la Ley de Benford. Así, las desviaciones de todos los dígitos se sumarían y el valor de este sumatorio es comparado con el valor crítico $X^2_{(\alpha, k-1)}$. La fórmula para el cálculo del estadístico es:

$$X^2 = \sum_{i=0}^9 \frac{(O_i - E_i)^2}{E_i} \quad \text{Ec. 4}$$

Donde O_i es la frecuencia observada que el dígito tome el valor i -ésimo.

E_i es la frecuencia esperada que el dígito tome el valor i -ésimo.

El Dr. Nigrini propone utilizar una u otra prueba de conformidad en función de si se examina la conformidad respecto a la distribución teórica de un dígito (prueba Z) o del conjunto de ellos (prueba x^2), el estadístico Z es un caso particular del X^2 para $n - 1$ grado de libertad. La primera permitiría detectar sobre apariciones de un dígito



concreto mientras que la segunda localizara patrones desajustados. Es de observar que las pruebas no son excluyentes sino complementarias; hallar desajuste en los datos respecto a los valores esperados nos haría sospechar del conjunto de registros.



El grafico 3 muestra el valor critico para una prueba unilateral de cola superior, para 8 grados de libertad, si el valor de chi-cuadrada calculada con la ecuación 4 es mayor a 15.51 podemos concluir que el conjunto de datos no se ajusta a una distribución de la ley de Benford.

Prueba de Arco de Mantis. (Mark J. Nigrini, 2012)⁵

Si los datos formaran un conjunto de Benford, las mantisas serían uniformes (distribuidas uniformemente) en el rango [0,1). No está claro cómo probar esta propiedad y cuánta tolerancia debe permitirse. Alexander⁹ (2009) propone una prueba rigurosa basada en las mantisas, que llamaremos prueba del arco de mantisa (MA).



En la cual, a cada mantisa se le asigna un rango desde 1 para el valor más pequeño hasta N para el valor más grande, El valor numérico de la mantisa se traza en el eje “y”. En la prueba arco mantisa, cada valor numérico se traza en un círculo unitario y un conjunto de Benford produce un conjunto de puntos distribuidos uniformemente en el círculo con radio 1 y centrado en el origen (0,0).

Si los datos difieren significativamente del centro de origen o también llamado centro de gravedad entonces no siguen La ley de Benford.

Para aplicar la prueba cada número debe convertirse en un punto en el círculo unitario. La x-coordenada y la y-coordenada de los puntos X_i se calculan usando las siguientes ecuaciones:

$$x - \text{coordenada} = \cos (2\pi * (\log_{10}(x_i) \bmod 1)) \text{ Ec. 5}$$

$$y - \text{coordenada} = \text{sen}(2\pi * (\log_{10}(x_i) \bmod 1)) \text{ Ec. 6}$$

Donde cos y sen se refieren a las funciones trigonométricas coseno y seno. El log, es el logaritmo de base 10, y mod1, es el módulo o parte fraccionaria del logaritmo (la parte a la derecha del punto decimal).

El centro de gravedad, también llamado vector medio y sus coordenadas en “x” y “y” se calcula mediante las siguientes ecuaciones:

$$\bar{x} - \text{coordenadas} = \frac{\sum_{i=1}^N \cos (2\pi * (\log_{10}(x_i) \bmod 1))}{N} \text{ Ec. 7}$$

$$\bar{y} - \text{coordenadas} = \frac{\sum_{i=1}^N \text{sen} (2\pi * (\log_{10}(x_i) \bmod 1))}{N} \text{ Ec. 8}$$

La longitud del vector medio, L^2 , se calcula usando la ecuación 9. El valor p (nivel de significancia) de L^2 se calcula usando la ecuación 10.

$$L^2 = (\bar{x} - \text{coordenadas})^2 + (\bar{y} - \text{coordenadas})^2 \text{ Ec. 9}$$

$$p - \text{valor} = e^{-L^2 N} \text{ Ec. 10}$$



La ecuación 9 muestra que a medida que el centro de gravedad se aleja cada vez más de (0,0), la longitud del vector medio aumenta. La ecuación 10 muestra que a medida que L^2 y N se hacen más grande (es decir que tienden al infinito), el valor de p tiende a 1. Como el vector medio tiende a 1 y N al infinito, el valor p tiende a 1. Un valor de p por encima de 0.05 es significativo al nivel de confianza de 0.05. el valor p se calcula según una distribución chi-cuadrada con 2 grados de libertad.



DISEÑO METODOLÓGICO

Tipo de estudio: Inferencial, Analítico, Retrospectivo

Inferencial porque se aplican pruebas estadísticas para validar la veracidad de los casos reportados

Analítico porque vamos a analizar en base a las pruebas estadísticas si la muestra aleatoria se ajusta a una distribución de la ley de Benford

Retrospectivo porque se estudió el primer semestre del año 2020.

De corte transversal: Se estudió los primeros seis meses del año 2020.

Área de estudio: Todos los países que la Universidad de Johns Hopkins reportó casos de COVID-19.

Población de estudio: 210 países que reportaron casos de COVID-19

Tamaño de Muestra: 14 países se calcularon a partir de la siguiente fórmula:

$$n = \frac{N * z_{\alpha/2}^2 * p * q}{e^2 * (N - 1) + z_{\alpha/2}^2 * p * q} \text{ Ec. 11}$$

En donde:

n = Tamaño de la muestra.

N = Tamaño de la población

Z = 1.96

e = Error de estimación máximo aceptado (25%).

P = Probabilidad de que un país sea seleccionado.

q = (1 - p) = Probabilidad de que un país no sea seleccionado.

α = Nivel de significancia (0.05)



Tipo de Muestreo: aleatorio simple probabilístico

Procedimiento de Muestreo

1. Se enumeraron los 210 países mediante un correlativo comprendidos entre 1 a 210.
2. Mediante la función “=ALEATORIO.ENTRE(inferior, superior)”, proporcionada por la aplicación Excel se generó 14 números aleatorios comprendidos entre 1 a 210.
3. Los números aleatorios obtenidos por esta función corresponden a los nombres de los países cuyo correlativo coincidan con estos números y que conformaron nuestra muestra de estudio.

Criterios de inclusión:

Países que se encuentran en la base de datos

Países que reportaron casos de COVID-19

Criterios de exclusión:

Países que no estén dentro de la base de datos

Países que no reportaron casos de COVID-19

Fuente de información:

- **Secundaria:** Base de datos de la universidad Johns Hopkins

Consideraciones éticas:

- Para mantener confidencialidad de los datos reportados, cada país estará codificado usando la letra “P” seguido de un número que va desde el 01 hasta el 14.
- No se revelará el nombre del país que incumplan la ley de Benford.



Método de recolección de datos: Descarga de la base de datos en línea en la página web de la **Universidad JHONS HOPKINS**.
<https://coronavirus.jhu.edu/map.html>

Plan de análisis: El procesamiento y análisis de los datos se realizó por medio de la hoja de cálculo Microsoft Excel 365.

Prueba Chi-Cuadrada: Prueba de hipótesis que compara la frecuencia observada de los datos con una distribución esperada y que analiza si un conjunto de datos se ajusta a una distribución teórica.

Hipótesis

H₀: Los dígitos se ajustan a la distribución de la Ley de Benford.

H₁: Los dígitos no se ajustan a la distribución de la Ley de Benford.

Nivel de significancia: $\alpha = 0.05$

Regiones críticas: Unilateral

Valor crítico: $\chi^2_{0.95,8} = 15.51$

Prueba Z: Prueba de hipótesis que compara la desviación de cada dato observados en relación con una distribución teórica.

Hipótesis

H₀: La proporción observada de un dígito no difiere significativamente de la proporción esperada según la distribución de la Ley de Benford.

H₁: La proporción observada de un dígito difiere significativamente de la proporción esperada según la distribución de Ley de Benford.

Nivel de significancia: $\alpha = 0.05$

Regiones críticas: Bilateral

Valor crítico: $Z_{0.975,8} = \pm 1.96$



Prueba Arco Mantis: Prueba de Hipótesis que pretende encontrar el centro de gravedad de un conjunto de dato de los cuales se esperaría que se distribuyan uniformemente en un círculo con radio 1 y centrado en el origen (0,0).

Hipótesis

H₀: La mantisa del conjunto de datos de COVID-19 se distribuye uniformemente alrededor de un círculo.

H₁: La mantisa del conjunto de datos de COVID-19 no se distribuye uniformemente alrededor de un círculo.

Nivel de significancia: $\alpha = 0.05$

Regiones criticas: Unilateral

Valor critico: $p - valor = e^{-L^2 N}$

Variables:

- Distribución de la Ley de Benford
- Números de casos reportados
- Número de veces que reportaron
- País
- Estadísticos de pruebas

Cruce de variables:

- Números de casos reportados Vs País
- Números de veces que reportaron Vs País
- Frecuencia porcentual observada de cada digito vs la frecuencia porcentual esperada de cada digito según la ley Benford
- País vs Estadísticos de pruebas



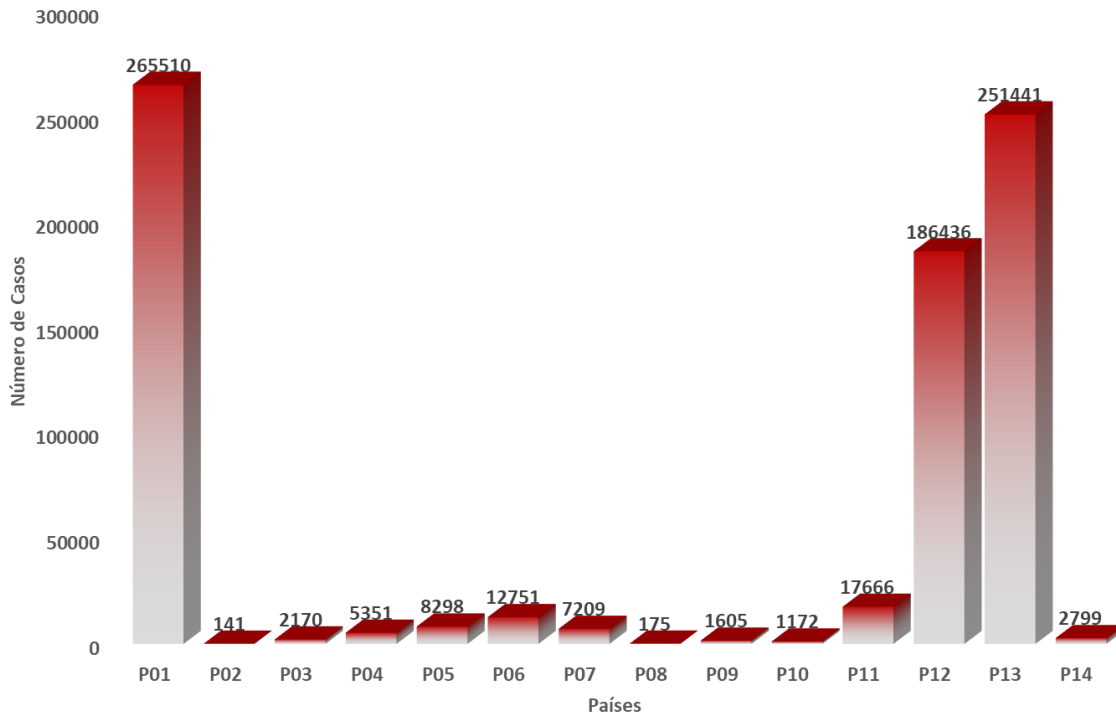
OPERACIONALIZACIÓN DE LAS VARIABLES

Variable	Conceptualización	Indicador	Escala de Medida
Distribución de la Ley de Benford	La probabilidad de ocurrencia de números es tal que las mantisas de sus logaritmos son equiprobables	Frecuencia de aparición de cada dígito	Razón
Número de casos reportados	Cifra de casos positivos de COVID-19 reportados por los países.	Casos Positivos de COVID-19	Discreta
País	Estado o nación políticamente independiente, con organización política y territorio libre soberano.	Bandera, Constitución política y territorio delimitado	Nominal
Número de veces que reportaron casos de Covid-19	Número de veces que un país reporto caso de COVID-19	Frecuencia de veces que un país reporto casos de COVID-19	Discreta
Estadísticos de pruebas	Prueba estadística que evalúa la evidencia de los datos proporcionados para probar una hipótesis.	Prueba Z: Prueba Chi-Cuadrada Prueba Arco Mantis	Continua o de intervalo



RESULTADOS Y DISCUSIÓN

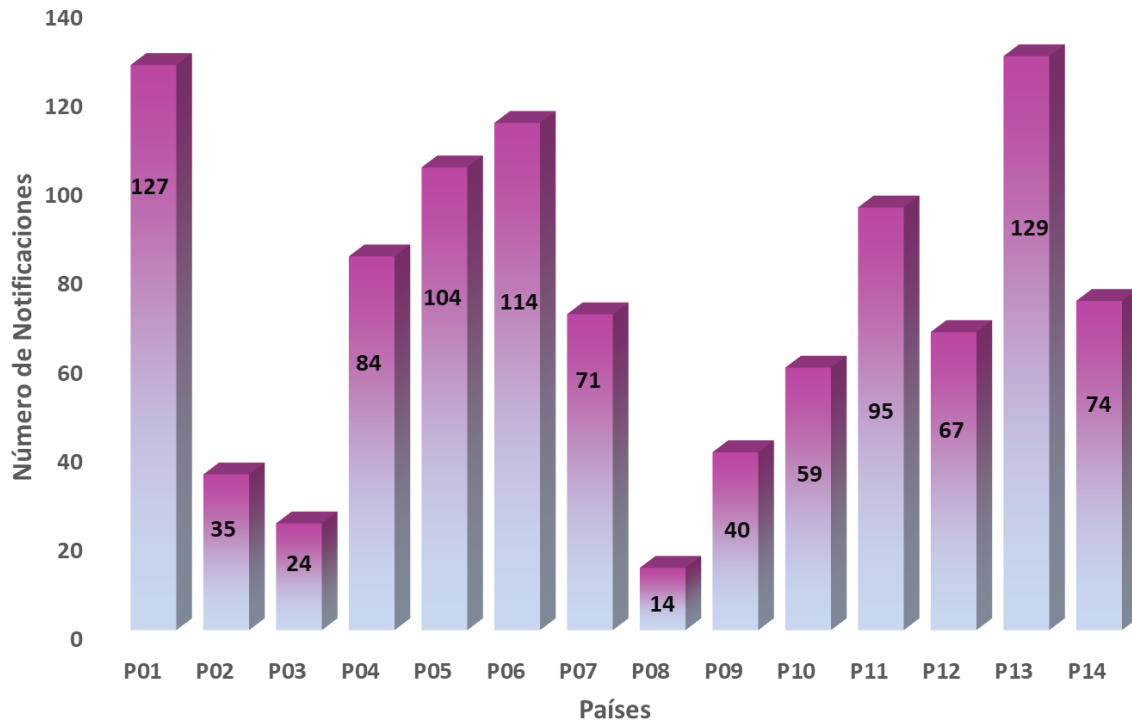
Gáfico 1. Número de Casos de COVID-19 Reportados por lo Países en Estudio en el Periodo Comprendido de Enero -Junio 2020



Los países P01, P12 y P13 reportaron mayores números de casos positivos de COVID-19, seguido de los países P04, P05, P06, P07 y P11 ambos grupos de países reportaron más de 5000 casos durante el periodo de estudio el resto de los países reportaron menos de esta cantidad.



Gráfico 2. Números de Notificaciones de COVID-19 Reportadas por los Países en Estudio en el Periodo Comprendido de Enero - Junio 2020

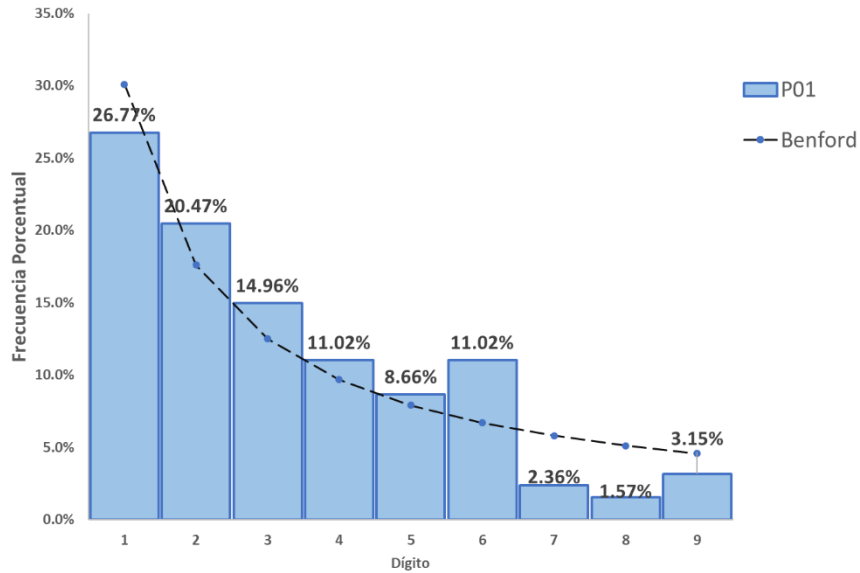


Los países P01, P05, P06 y P13 realizaron el mayor número de reporte de casos positivos de COVID-19, el resto de los países reportaron con menor frecuencia durante el periodo de estudio. Hay que aclarar que el número de reporte no es un factor determinante del cumplimiento o incumplimiento de la ley de Benford ya que se ha comprobado que países que realizan un número menor de reportes pueden seguir una mejor distribución de la ley de Benford.



Gráfico 3. Distribución de frecuencia del Primer Dígito significativos de los Números de casos de COVID-19 Reportados por los países en estudio en el periodo comprendido de enero - junio 2020

País P01



País P02

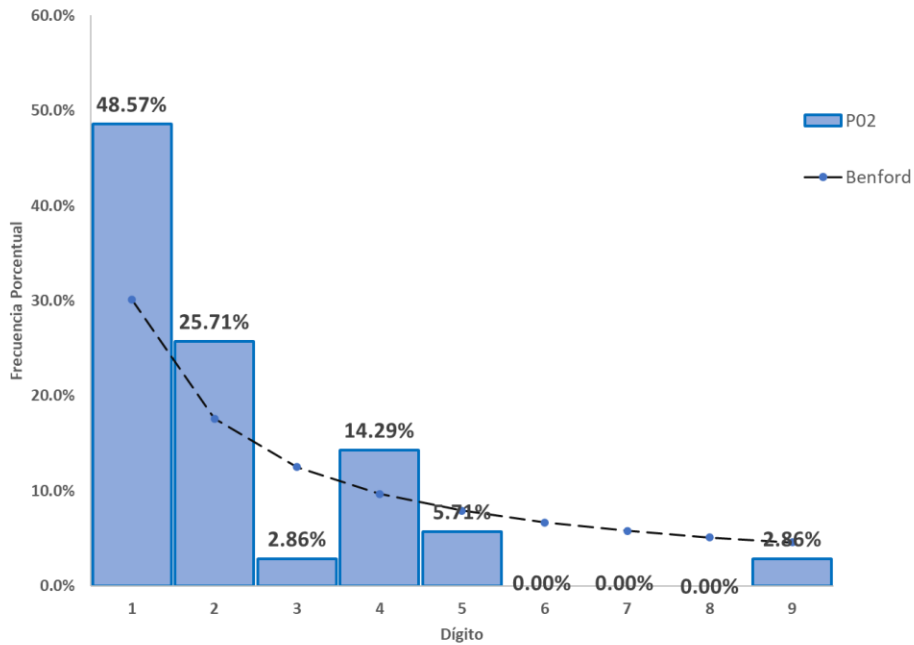
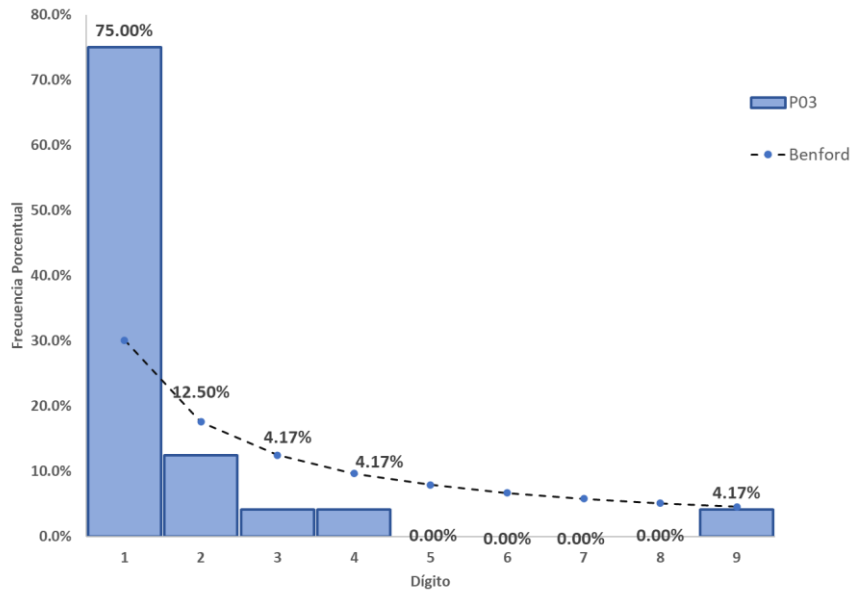




Gráfico 3. Continuación

País P03



País P04

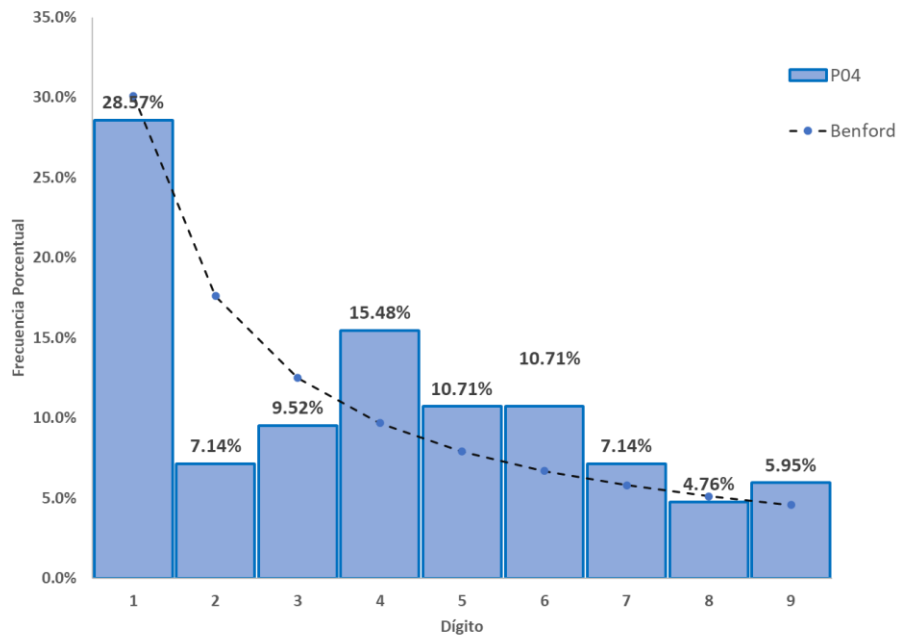
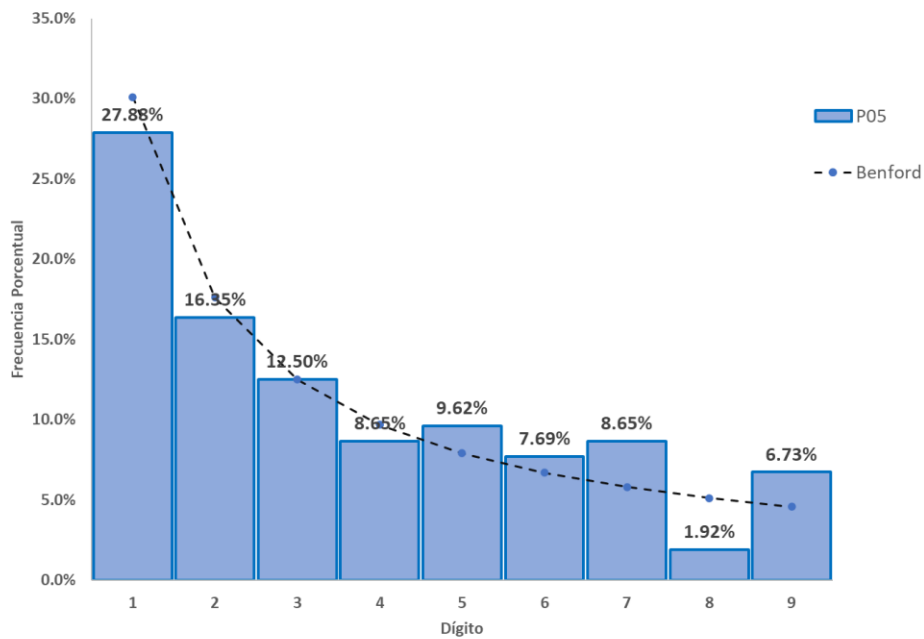




Gráfico 3. Continuación

País P05



País P06

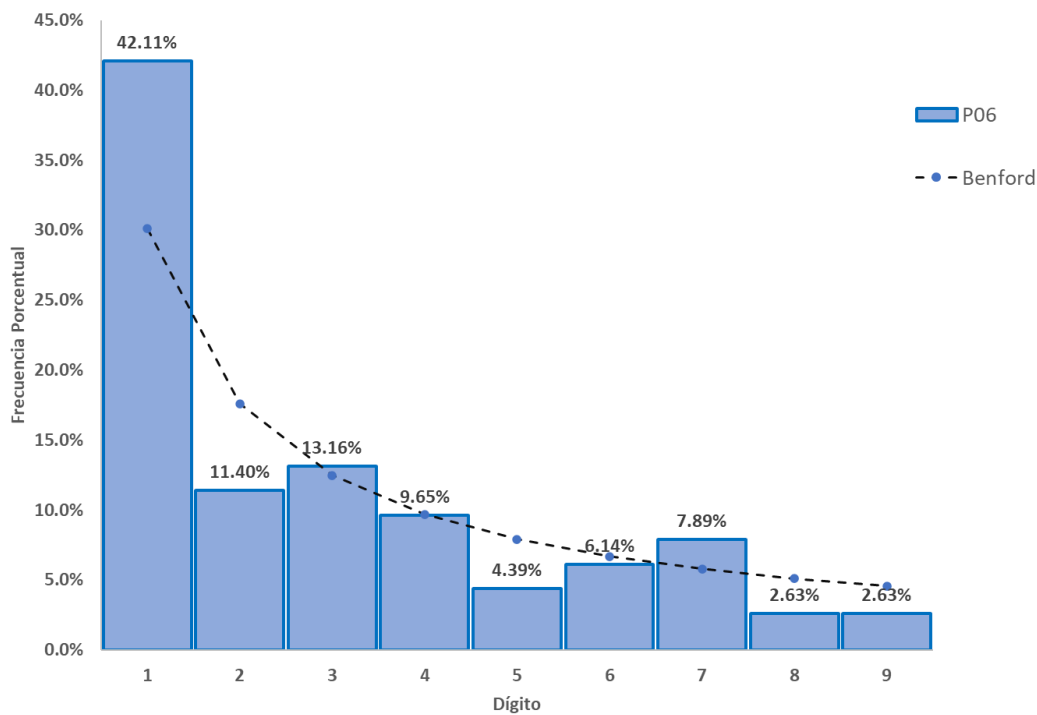
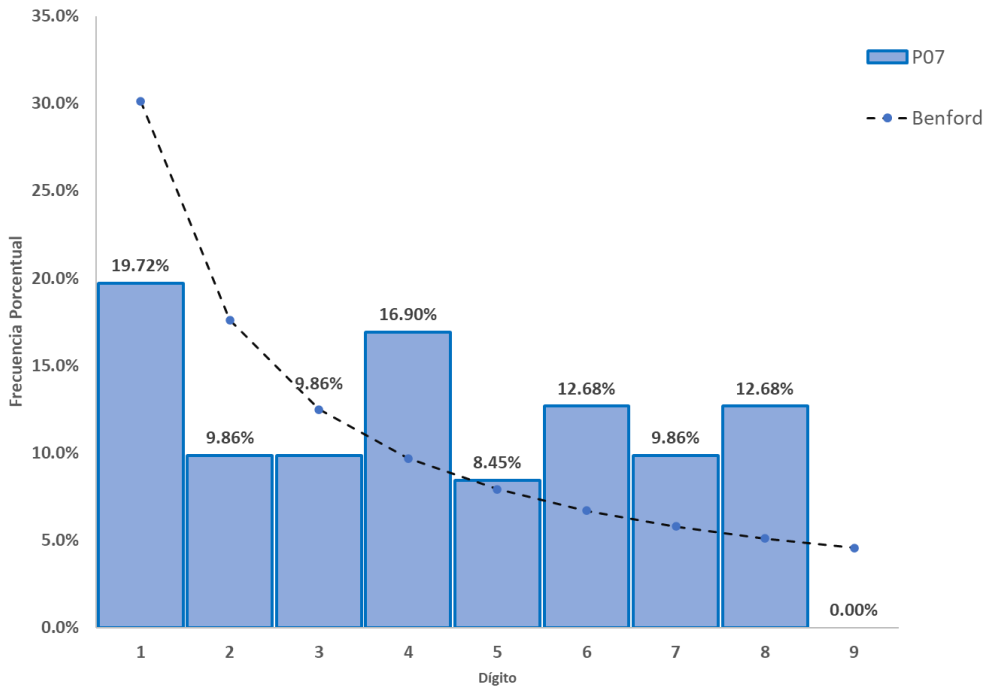




Gráfico 3. Continuación

País P07



País P08

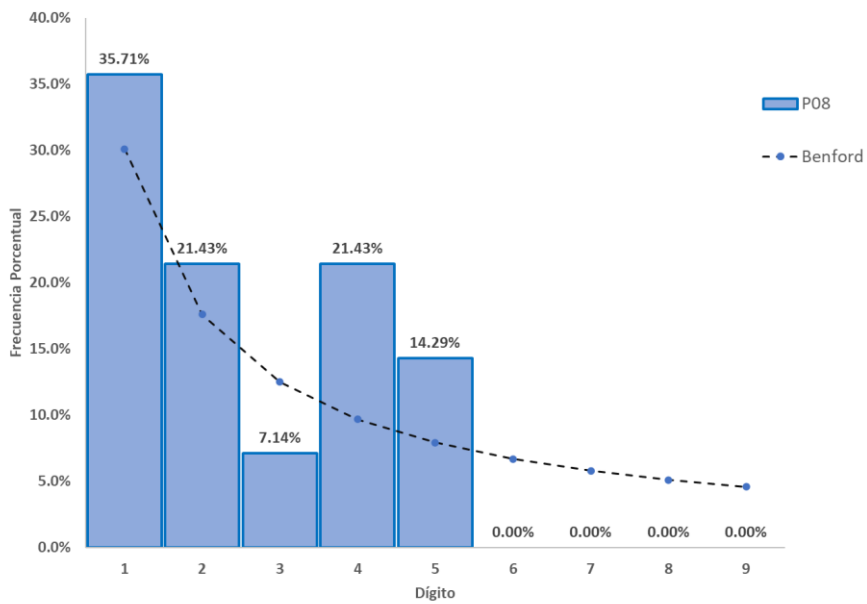
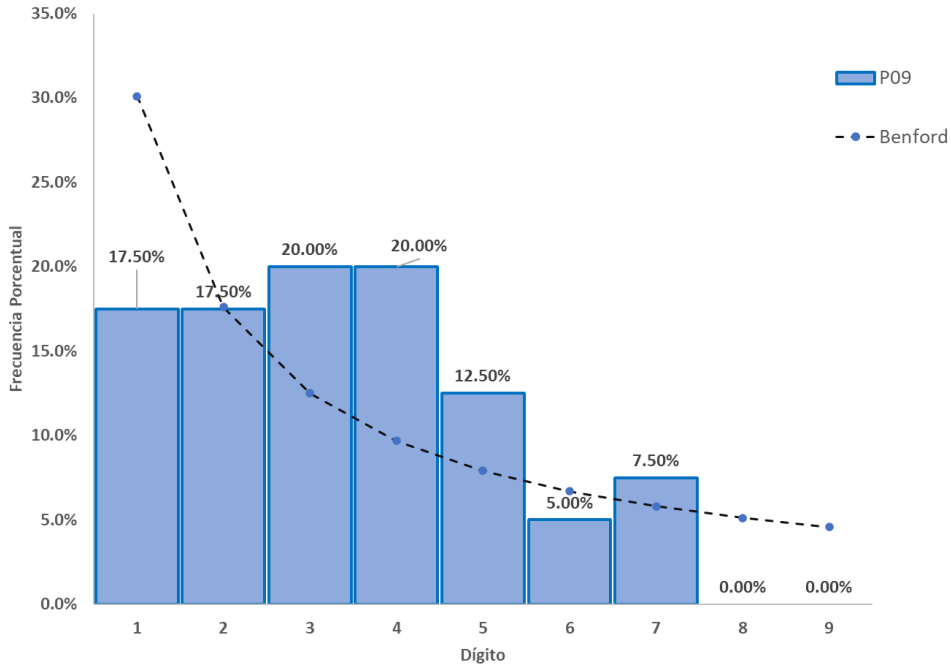




Gráfico 3. Continuación

País 09



País P10

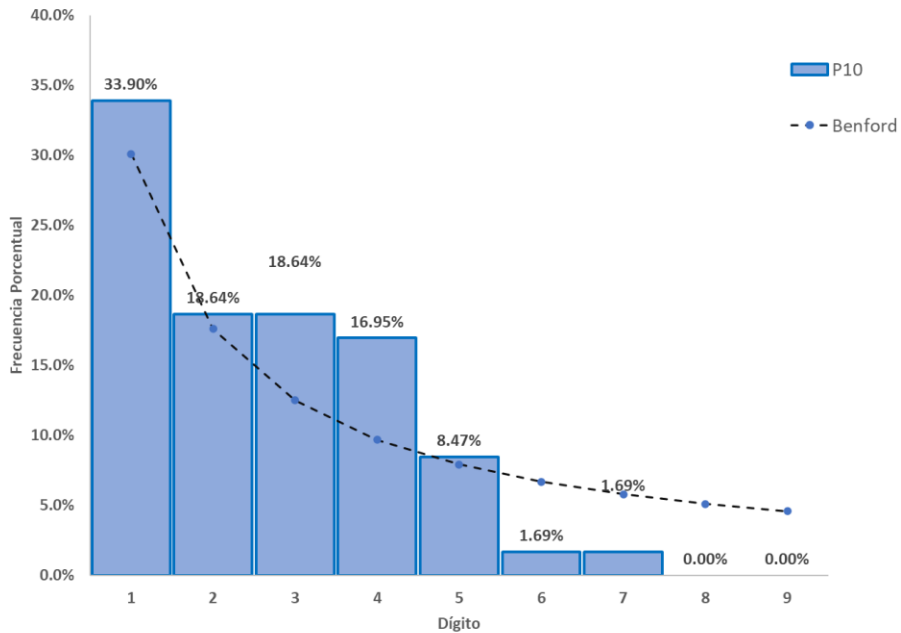
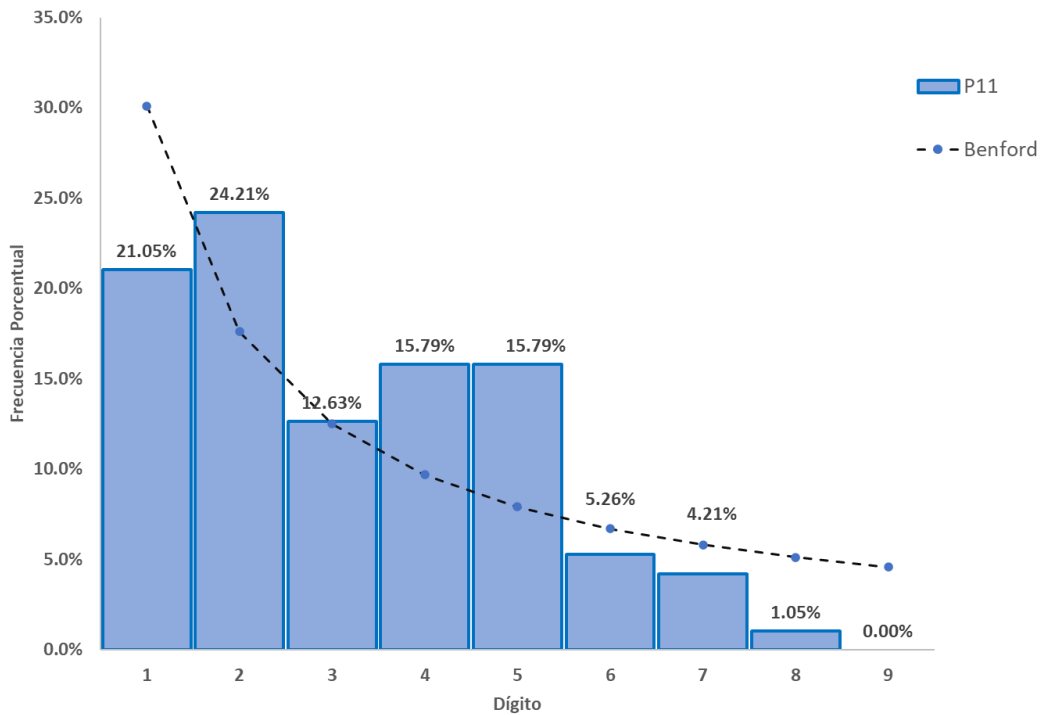




Gráfico 3. Continuación

País P11



País P12

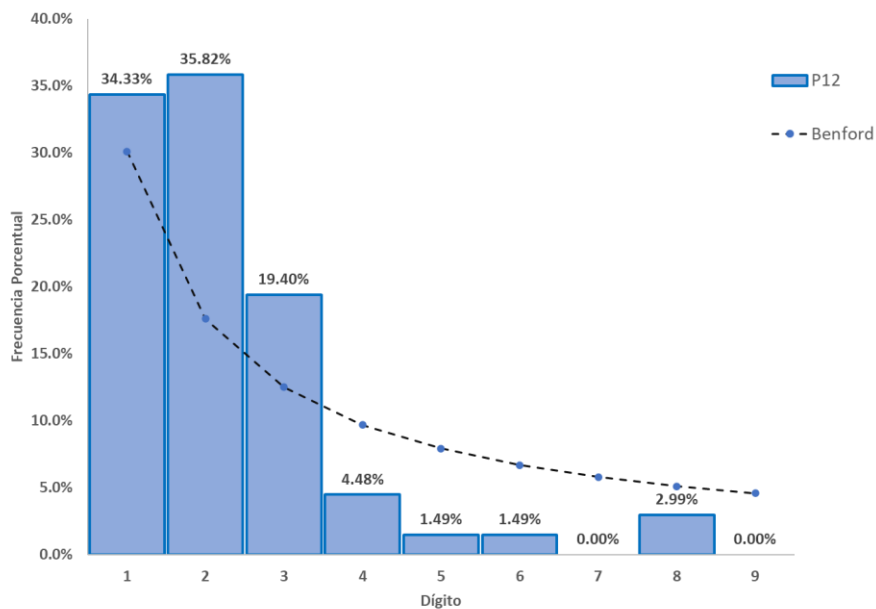
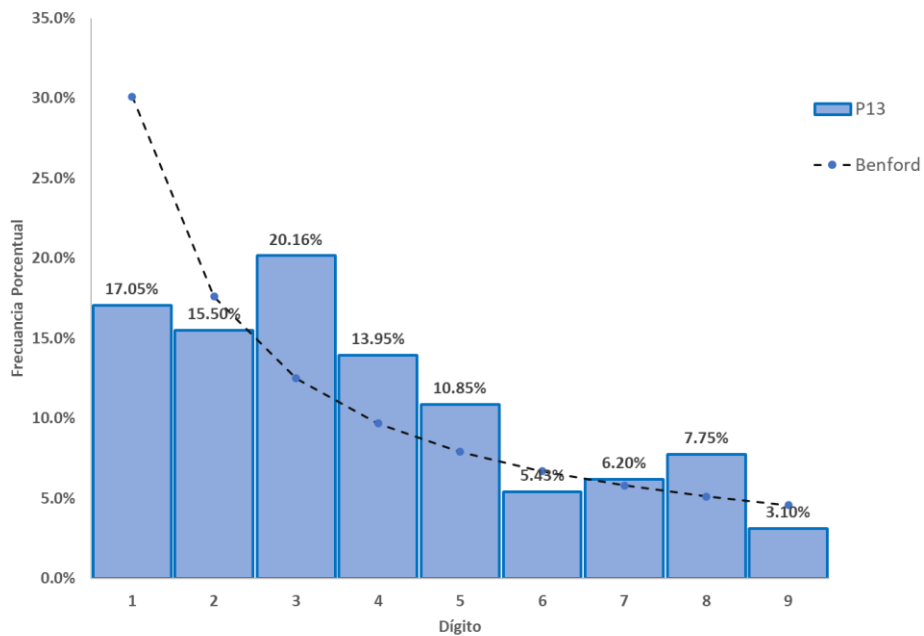


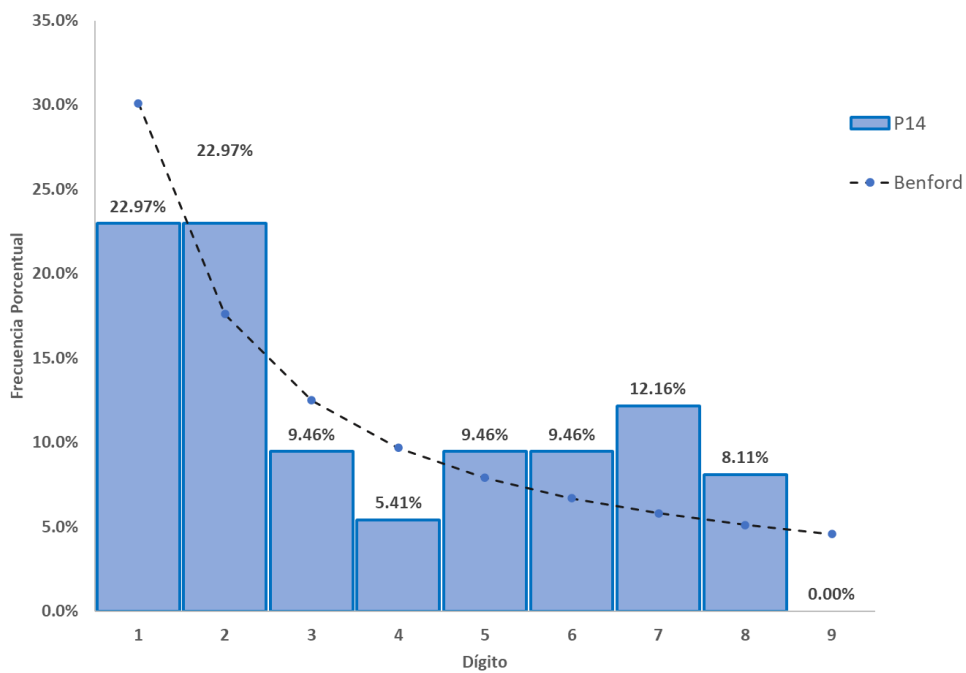


Gráfico 3. Continuación

País P13



País P14





De manera visual podemos observar las desviaciones, de cada dígito (grafico de columna), con respecto a la distribución de Benford (grafico de línea discontinua), para los 14 países en estudio. De los 14 países solo los países P01 y P05 dan evidencia observacional que sus datos se ajustan a la distribución de la ley de Benford, podemos notar que la línea discontinua para dichos gráficos sigue un comportamiento logarítmico, es decir la frecuencia porcentual observadas para cada dígitos va decreciendo a medida que decrece la frecuencia porcentual esperada, otra evidencia que confirma que los países P01 y P05 siguen una distribución según la ley de Benford, es la forma de los gráficos de barra, podremos notar que la altura de la barra tiende a decrecer a medida que incrementa el digito (1 a 9).

.



Tabla 1. Estadístico Z para el Primer Dígito significativos de los Números de casos de COVID-19 Reportados por los países en estudio en el periodo comprendido de enero - junio 2020

Dígito \ País	1	2	3	4	5	6	7	8	9
P01	0.722	0.731	0.707	0.358	0.146	1.774	1.467	1.610	0.557
P02	2.198	1.037	1.469	0.633	0.170	1.247	1.106	0.990	0.082
P03	4.573	0.389	0.925	0.570	1.059	0.904	0.779	0.674	-0.393
P04	0.187	2.375	0.658	1.608	0.747	1.256	0.293	-0.101	0.343
P05	0.386	0.232	0.674	0.250	1.335	1.025	1.947	0.728	1.648
P06	2.692	0.958	0.954	0.682	0.537	0.335	1.661	0.389	0.163
P07	1.778	1.559	0.492	1.853	-0.054	1.779	1.210	2.622	1.561
P08	0.166	0.024	0.201	1.033	0.387	0.468	0.357	0.262	0.180
P09	1.565	-0.189	1.197	1.937	0.780	0.113	0.122	1.110	1.007
P10	0.494	0.038	1.232	1.664	-0.083	1.276	1.070	1.488	1.370
P11	1.811	1.555	-0.115	1.836	2.651	0.353	0.443	1.565	1.889
P12	0.621	3.753	1.526	1.236	1.722	1.459	1.770	0.514	1.500
P13	3.135	0.512	2.498	1.488	1.071	0.400	0.007	1.159	0.591
P14	1.210	1.059	0.614	1.050	0.276	0.719	2.093	0.905	1.606

Los países P02, P03, P06 y P13 difieren en el dígito 1 en proporción a los esperados según la ley de Benford, los países P04 y P12 lo hacen en relación con el dígito 2, el país P13 lo hace con respecto al dígito 3, el P11 lo hace en dependencia al dígito 5, P14 en relación con el dígito 7 y el P07 lo hace con respecto al dígito 8. Esto nos indica que debemos revisar todos los registros de casos de COVID-19 que reportaron estos países y que inician con cada uno de estos dígitos para analizar por qué se desvían de la proporción esperada según Benford.



Tabla 2. Estadístico Chi-Cuadrado y arco mantisa para el Primer Dígito significativos de los Números de casos de COVID-19 Reportados por los países en estudio en el periodo comprendido de enero - junio 2020

Prueba País	Chi-Cuadrada		Arco Mantis	
	X ²	P-valor	L ²	P-valor
P01	11.818	0.160	0.012	0.211
P02	15.239	0.055	0.077	0.069
P03	24.650	0.002	0.268	0.002
P04	12.272	0.139	0.066	0.004
P05	13.891	0.085	0.003	0.741
P06	13.033	0.111	0.008	0.402
P07	26.190	0.001	0.043	0.047
P08	6.396	0.603	0.007	0.904
P09	13.609	0.093	0.039	0.212
P10	14.969	0.060	0.0336	0.137
P11	24.133	0.002	0.0742	0.001
P12	31.203	0.000	0.101	0.001
P13	20.215	0.010	0.055	0.113
P14	15.322	0.053	0.010	0.474

Los datos reportados de COVID-19 por los países P03, P07, P011 y P12 no se ajustan a la ley Benford según la prueba estadística X² y Arco mantisa. El país P04 cumple con el estadístico de prueba X² y no cumple con la de Arco mantisa, El país P13 no cumple con la prueba estadístico X² y cumple con la de Arco mantisa, aunque estos dos países cumplen con una prueba y/o incumple con la otra, no podemos afirmar, que no se ajustan a ley, puesto que se requeriría un análisis más profundo para obtener un juicio más acertado. Al mismo tiempo podemos notar que los países que no se ajustan a la ley, mediante estas pruebas, también difiere significativamente de la proporción esperada de un dígito según el estadístico Z, exceptuando los países P01, P02 P05, P06, P08, P09, P10 y P14 que cumplen ambas pruebas estadísticas por lo cual se ajusta a la Ley de Benford.



CONCLUSIONES

- ✓ Las evidencia obtenida en esta investigación no constituyen prueba de invalidez de los datos de COVID-19 reportados por los países en estudio, pero hemos identificados, mediante el estadístico Z, que 10 países muestra desviaciones de un dígito de la ley de Benford, principalmente en reporte de casos que inician con los dígitos 1, 2, 3, 5, 7 y 8. Al mismo tiempo, los estadísticos de prueba, chi-cuadrada y arco mantisa evidenciaron que la distribución de probabilidad de 8 países se ajustan perfectamente, para todos los dígitos, a una distribución de Benford, indicio suficiente para asegurar que estos 8 países reportaron con veracidad sus casos de COVID-19 durante el periodo de estudio.
- ✓ Cabe resaltar que con este procedimiento no se puede afirmar, de ninguna manera, que los países en estudio estén cometiendo fraude; pero si es útil como indicador o alerta; que se está ante un comportamiento anómalo en estas cuentas, lo que evidencia la importancia y la necesidad de realizar revisiones exhaustivas a los informes de COVID-19 reportados.



RECOMENDACIONES

De manera General:

Extender el estudio expuesto a otras áreas del conocimiento de la salud tales como:

- Estudios epidemiológicos transmitidos por vectores como: dengue, malaria, zika, etc. Tasas de morbi-mortalidad y en la administración y gerencia farmacéutica
- Capacitar a todos los profesionales sobre los pasos, de este tipo de herramientas estadística para ayudar al manejo correcto de los datos obtenidos en sus investigaciones
- Apoyar a los ministerios de salud de manera general, en la mejora de los procesos de mediciones estadística, lo cual garantizara datos confiables y de cierta manera apoyara a una mejor en calidad de vida.



BIBLIOGRAFÍA

1. Newcomb Simon (1881). Note on the Frequency of Use of the Different Digits in Natural Numbers. *American Journal of Mathematics*, 4(1), 39 – 40.
2. Benford Frank (1938). The Law Anomalous Numbre, *Proceedings of the American Philosophical Society*, 78(4), 551 – 572.
3. Pinkham Roger S. (1961). On the Distribution of First Significant Digits. *The Annals of Mathematical Statistics*, 32(4), 1223 – 1230.
4. Nigrini Mark J. (1994). Using Digital Frequencies to Detect Fraud, *The White Paper*, April/May, pp. 3 – 6.
5. Nigrini Mark J. (2012). *Benford's Law Applications for Forensic Accounting, Auditing, and Fraud Detection*. John Wiley & Sons, Inc.
6. Hurtado Alexandra. (2020). Evaluación Comparativa de la Vigilancia en Salud Pública de COVID-19 en Colombia: Primer Semestre, *Biomédica*, 40(2), 131 – 138.
7. Isea Raúl (2020). La Dinámica de Transmisión del COVID-19 desde una Perspectiva Matemática, *OC*. 5(1), 15 – 23.
8. Martínez G. Raúl A. Canizales R. Carlos E. (2009). Ley de Benford y sus Aplicaciones. <http://ri.ues.edu.sv/id/eprint/12497/1/19200765.pdf>
9. Alexander, J. (2009). Remarks on the use of Benford's Law. Working Paper, Case Western Reserve University, Department of Mathematics and Cognitive Science.