Impacto del COVID-19 en el número de ciberataques en las regiones de Europa y Estados Unidos, en contraste con el periodo prepandémico 2016-2019.

Jose Ricardo Acuña González [1], Johan Castaño Bustamante[1]. jose.acunagonzalez@ucr.ac.cr
johan.castano@ucr.ac.cr

Resumen

Esta investigación examina el impacto del COVID-19 en el incremento de los ciberataques en Europa y Estados Unidos durante el período 2016-2023, empleando pruebas estadísticas clave como Kolmogórov-Smirnov, exacta de Fisher, prueba de signos y Brown-Forsythe. Durante la pandemia, la rápida transición hacia el trabajo remoto y la mayor dependencia de servicios digitales crearon un entorno propicio para los ciberdelincuentes, quienes aprovecharon las vulnerabilidades expuestas. El análisis con Kolmogórov-Smirnov busca identificar cambios en la distribución de los ataques cibernéticos, así como determinar la normalidad de las muestras. La prueba exacta de Fisher se utiliza para determinar la relación entre las muestras. Además, la prueba de signos evalúa diferencias en las medianas de ataques antes y después de la pandemia, destacando un aumento significativo en la frecuencia de estos. Lo anterior, sumado a la prueba Brown-Forsythe, logra determinar una diferencia en el comportamiento de los ciberataques del periodo 2020-2023 comparado al 2016-2019. Mediante el análisis se pretende generar conciencia sobre la importancia de mantener sistemas de seguridad robustos, tanto para empresas, gobierno y ciudadanos.

Palabras Claves: Ciberataque, Región, Impacto del COVID-19,Pandemia.

INTRODUCCIÓN

El presente proyecto tiene como objetivo determinar la existencia de un cambio en el número de ciberataques en las regiones de Europa y Estados Unidos durante el periodo de la pandemia de COVID-19 (2020-2023) en comparación con el periodo pre-COVID (2016-2019). Es importante destacar que la pandemia de COVID-19 ha transformado radicalmente numerosos aspectos de la vida cotidiana y profesional, provocando una aceleración sin precedentes en la adopción de tecnologías digitales. Cada año que pasa, las tecnologías toman más y más protagonismo en el día a día de los ciudadanos, las instituciones gubernamentales y privadas. La exposición a riesgos, en este entorno virtual, es latente, por lo que el estudio de esta materia se torna fundamental para la concientización de un mundo digital donde la ciberseguridad sea una pieza angular del desarrollo. Mediante un análisis de la literatura científica, se inicia el desarrollo del estudio del impacto del Covid-19 en el número de ciberataques en Estados Unidos y Europa. Los autores Lallie y cols. (2021) proporcionan un análisis detallado del incremento en los ciberataques durante la pandemia. Los autores presentan una cronología de eventos cibernéticos desde el inicio de la crisis sanitaria, destacando que el volumen y la sofisticación de los ataques cibernéticos aumentaron significativamente. La transición masiva al teletrabajo y el mayor uso de servicios en línea crearon un entorno favorable para los ciberdelincuentes, quienes aprovecharon las vulnerabilidades en las infraestructuras de seguridad de las empresas.

Asimismo, Wiggen (2020) analiza el impacto del COVID-19 no solo en el cibercrimen convencional, sino también en las actividades cibernéticas patrocinadas por el estado. Durante la pandemia, se observó un aumento en las actividades de espionaje cibernético y los ciberataques dirigidos a infraestructuras críticas, con estados-nación aprovechando la distracción global para avanzar sus agendas estratégicas. El informe de Wiggen subraya que tanto Estados Unidos como países europeos fueron objetivos de campañas de desinformación y ataques cibernéticos que buscaban desestabilizar sus sistemas políticos y económicos. Las tácticas empleadas incluyeron la propagación de noticias falsas sobre la COVID-19 y ataques a las cadenas de suministro de vacunas, con el objetivo de socavar la confianza pública y crear caos.

[1] Estudiante de Ciencias Actuariales en la Universidad de Costa Rica

Yadav y cols. (2021) abordan las diversas amenazas de ciberseguridad que emergieron durante la pandemia, destacando que el rápido cambio hacia el teletrabajo expuso a muchas organizaciones a nuevos riesgos. La falta de preparación para un entorno de trabajo remoto y la utilización de redes domésticas inseguras facilitaron la actividad de los ciberdelincuentes. El estudio identifica varias formas de ciberataques que se intensificaron durante este periodo, incluyendo el aumento en los ataques de denegación de servicio (DDoS), ataques de ingeniería social y el uso de malware. Yadav y cols. (2021) también discuten la importancia de fortalecer las políticas de ciberseguridad e invertir en tecnologías que permitan la detección y mitigación de amenazas en tiempo real.

Además, el estudio realizado por Hawdon, Parti, y Dearden (2020) caracteriza los tipos de ciberataques tanto en el periodo pre-COVID-19 como en el periodo post-COVID-19. Utilizando la prueba de chi-cuadrado (X^2), Hawdon busca determinar si existe un cambio significativo entre ambos periodos, especialmente en el periodo post-COVID-19 en comparación con el periodo pre-COVID-19. Además, la autora emplea la prueba T de Student (T-test) en ambos periodos para evaluar si las medias de los diferentes tipos de ciberataques muestran cambios significativos.

Por ultimo, el impacto de la pandemia de COVID-19 en la ciberseguridad ha sido profundo y polifacético. La rápida adopción de tecnologías digitales y el cambio a entornos de trabajo remoto ampliaron la superficie de ataque para los ciberdelincuentes y los actores estatales maliciosos. Los estudios revisados muestran un aumento significativo en la frecuencia y sofisticación de los ciberataques, con un enfoque particular en phishing, ransomware y espionaje cibernético. Es crucial que tanto las organizaciones como los gobiernos inviertan en medidas de ciberseguridad más robustas y se mantengan vigilantes ante las amenazas en evolución. La pandemia ha dejado en claro que la ciberseguridad debe ser una prioridad central en la era digital para proteger la integridad y la resiliencia de las infraestructuras críticas.

Ahora, para el desarrollo de la investigación es de importancia contar con pruebas estadísticas de detección de cambios en los datos, ya que de esta manera, se lograría evidenciar que existió una diferencia entre los 2 periodos de tiempo. En este caso, la línea divisora y temporal entre los dos grupos corresponde a la pandemia del COVID-19. Por tanto, se decide realizar las siguientes pruebas que ayudaran a identificar esta diferencia, robustecer el análisis y poder fundamentar la investigación con herramientas estadísticas relevantes.

METODOLOGÍA

A continuación, se presenta una descripción detallada de los datos, los cuales serán utilizados para realizar pruebas estadísticas y responder a la pregunta del tema:

- Fuente de información: Repositorio Europeo de Ciber incidentes (EUREPOC, 01/04/2024) (Universidad de Heidelberg, Universidad de Innsbruck, Fundación de Ciencias y Política de Alemania y el Instituto de Ciber Policía de Estonia).

- Contexto temporal y espacial de los datos: 01/01/2000 hasta 01/04/2024 dentro de las regiones de Norteamérica, Centroamérica, Antillas, Sudamérica, Europa Occidental, Europa Oriental, Cáucaso, Siberia, Asia Central, Asia Oriental, África del Norte, África subsahariana, Cercano/Medio Oriente, Islas del Pacífico, Australia.

- Facilidad de obtener la información: Alta. El repositorio presenta una fácil accesibilidad a los datos y una amplia exposición de estos.

- Población de estudio: Ciberataques registrados dentro del repositorio europeo de ciber intendentes con las regiones de Norteamérica, Centroamérica, Antillas, Sudamérica, Europa Occidental, Europa Oriental, Cáucaso,

Siberia, Asia Central, Asia Oriental, África del Norte, África subsahariana, Cercano/Medio Oriente, Islas del Pacífico, Australia como receptores de ciberataques del 2000-2024.

- Muestra observada: Los 2789 ciberataques registrados dentro del repositorio europeo de ciber incidentes para las regiones de Norteamérica, Centroamérica, Antillas, Sudamérica, Europa Occidental, Europa Oriental, Cáucaso, Siberia, Asia Central, Asia Oriental, África del Norte, África subsahariana, Cercano/Medio Oriente, Islas del Pacífico, Australia desde 2000 hasta 2024.

- Unidad estadística o individuos: Cada ciberataque registrado dentro del repositorio europeo de ciber incidentes para las regiones de Norteamérica, Centroamérica, Antillas, Sudamérica, Europa Occidental, Europa Oriental, Cáucaso, Siberia, Asia Central, Asia Oriental, África del Norte, África subsahariana, Cercano/Medio Oriente, Islas del Pacífico, Australia. Desde 2000 hasta 2024 y sus variables.

- Descripción de las variables de la tabla: La base de datos original presenta 20 variables, de las cuales para el desarrollo de esta investigación, solo se consideraron 5, esto debido a que de las 15 variables adicionales, 5 no eran de interés para esta investigación y con respecto las 10 restantes, debido a la cantidad de datos faltantes y a la poca cantidad de observaciones con respecto a las regiones de interés, no se consideraron, por lo cual las variables a considerar son:

    - start date: Fecha de inicio del ciberataque
    - receiver country: País que recibió el ciberataque
    - receiver region: Región que recibió el ciberataque
    - target multiplier: Importancia del ciberataque, ya sea moderada,
    - impact indicator value: Valor numérico relacionado con el indicador de impacto, con valores entre 0 y 13

Una muestra de la base con las variables de interés corresponde a:

Cuadro 1: Muestra de la base de datos con las variables de interes

start_date	receiver_country	receiver_region	target_multiplier	impact_indicator_value
2023-12-30	United States	NATO; NORTHAM	Moderate - high political importance	0
2024-03-17	Italy	EUROPE; NATO; EU(MS)	Moderate - high political importance	0
2022-01-01	Jordan	ASIA	Moderate - high political importance	0
2024-03-15	Australia	OC	Moderate - high political importance	0
2024-03-01	United Kingdom	EUROPE; NATO; NORTHEU	Moderate - high political importance	0
2024-03-14	United States	NATO; NORTHAM	Moderate - high political importance	0

Fuente: Elaboración propia con datos del Repositorio Europeo de Ciberincidentes

Razón de ciberataques por usuario de internet

Consideran que el objetivo de la investigación es determinar la existencia de un impacto en el numero de ciberataques en 2 periodos de tiempo, se determina la necesidad de comparar los 2 periodos de tiempo, con lo cual se

necesita que la muestra de datos de ciberataques este en valores reales. Para esto, se decidió establecer una razón o proporción de los datos en relación con la cantidad de usuarios que utilizan internet. A partir de los datos presentes en la plataforma de Statista statista (2024a) y statista (2024b), una empresa que desarrolla bases de datos confiables, se encontraron los datos relacionados con la cantidad de usuarios de internet para Europa y Estados Unidos en millones.

A partir de los datos anuales sobre el número de usuarios de internet en Estados Unidos y Europa entre 2016 y 2023, y considerando la base de datos disponible, se decide calcular una razón o proporción cuatrimestral. Dado que los usuarios de internet no tienden a cambiar tan significativamente durante un año y considerando la falta de más bases de datos, se opta por dividir la cantidad de usuarios de internet de cada año en 3 partes, representando así los valores cuatrimestrales.

Este enfoque permite determinar tres valores cuatrimestrales por año con los datos anuales disponibles. Luego, se calcula la proporción entre la cantidad de ciberataques cuatrimestrales y el número de usuarios de internet en cada cuatrimestre, para las regiones de interés durante el período 2016-2023. De esta manera, se establece una razón o proporción cuatrimestral de los datos, que será útil para futuros cálculos y la aplicación de pruebas estadísticas.

Por ultimo, debido a la cantidad de observaciones reducida se determina que para lo referente a esta investigación, se considerara un nivel de significancia del 0.15

Pruebas Estadísticas

Prueba de Kolmogórov-Smirnov

Según DeGroot (2002), La prueba de Kolmogórov-Smirnov se utiliza para determinar si dos conjuntos de datos tienen la misma distribución, esto mediante la comparación de la función de distribución acumulada empírica de los datos muéstrales con respecto a la distribución esperada, con lo cual se define:
Hipótesis nula

$$H_0 : f(X) = f^*(x)$$

y la hipótesis alternativa

$$H_1 : f(X) \neq f^*(x)$$

con f(x) la función de distribución desconocida asociada a un conjunto de observaciones $X_1, X_2, ..., X_n$ y $f^*(x)$ es la función de distribución desconocida asociada a un conjunto observaciones $Y_1, Y_2, ..., Y_m$.

Con $f_n(x)$ la función de distribución calculada a partir de los valores $X_1, ..., X_n$ y $f_m^*(x)$ la función de distribución calculada a partir de los valores de $Y_1, ..., Y_m$. de esta manera se define el estadístico D_{nm} que representa la máxima diferencia entre la función de distribución acumulada (c.d.f) de la muestra observada y la teórica:

$$D_{nm} = \sup_{x \in R} = [f_n(X) - f_m^*(x)]$$

Si $D_{nm} \rightarrow 0$ cuando n,m $\rightarrow \infty$ entonces H_0: es verdadera

Prueba de Signos

Según Hollander, M., Wolfe, D. A. (1999) la prueba de signos es una prueba no paramétrica utilizada para evaluar la hipótesis nula de que las medianas de dos distribuciones emparejadas son iguales. Es una alternativa a la prueba t de muestras pareadas cuando no se puede asumir que los datos siguen una distribución normal. Esta prueba es útil especialmente cuando los datos son ordinales o cuando las suposiciones de normalidad no se cumplen.

Para el procedimiento de la primera primero se calculan las diferencias $D_i = X_i - Y_i$ para cada par de observaciones. En segundo lugar se cuenta el número de diferencias positivas (S^+) y negativas (S^-). Luego se calcula el estadístico de prueba el cual corresponde al menor de los dos conteos ($S = \min(S^+, S^-)$). Bajo la hipótesis nula, S sigue una distribución binomial con parámetros n y $p = 0{,}5$, donde n es el número total de diferencias no nulas. Por último se calcula el valor p usando la distribución binomial:

$$p = 2 \sum_{k=0}^{S} \binom{n}{k} \left(\frac{1}{2}\right)^n$$

Prueba Brown-Forsythe para dos muestras

Según Brown MB (1974) la prueba de Brown-Forsythe es una prueba estadística utilizada para evaluar la igualdad de varianzas entre dos o más grupos. Es una modificación de la prueba de Levene que utiliza la mediana en lugar de la media, haciéndola menos sensible a distribuciones no normales.

Formulación de la Prueba

Hipótesis Nula (H_0): Las varianzas de las poblaciones son iguales.

Hipótesis Alternativa (H_1): Las varianzas son diferentes.

El estadístico W para la prueba de Brown-Forsythe se define como:

$$W = \frac{(N-k)}{(k-1)} \cdot \frac{\sum_{i=1}^{k} N_i (\bar{Y}_i - \bar{Y})^2}{\sum_{i=1}^{k} \sum_{j=1}^{N_i} (Y_{ij} - \bar{Y}_i)^2}$$

donde: N es el tamaño total de la muestra, k es el número de grupos, N_i es el tamaño de la muestra del grupo i, $\bar{Y}_i$ es la media de las desviaciones absolutas respecto a la mediana del grupo i, $\bar{Y}$ es la media global de las desviaciones absolutas, Y_{ij} es la desviación absoluta de la j-ésima observación respecto a la mediana del grupo i.

Prueba Exacta de Fisher

Según Mehta CR (1983), la prueba exacta de Fisher es una técnica estadística utilizada para evaluar la significancia de la asociación entre dos variables categóricas en una tabla de contingencia $2x2$. Es especialmente útil para muestras pequeñas y cuando los datos no cumplen con los requisitos de la prueba chi-cuadrado.

Formulación de la prueba:

Hipótesis Nula (H_0): Las dos variables categóricas son independientes; no existe una asociación significativa entre ellas.

Hipótesis Alternativa (H_1): Las dos variables categóricas no son independientes; existe una asociación significativa entre ellas.

Primeramente, hay que organizar los datos en una tabla de contingencia 2×2 y según DeGroot (2002) las tablas de contingencia son una herramienta apropiada para analizar la relación entre dos variables categóricas, donde está se define como un arreglo bidimensional en el que cada observación se puede clasificar de dos o más formas, generalmente a lo largo de filas y columnas. El arreglo contiene la siguiente construcción:

R representa el número de filas en la tabla.

C representa el número de columnas en la tabla.

N_{ij} representa el número de individuos en la muestra clasificados en la fila i y columna j.

N_{i+}representa el total de individuos en la fila i , calculado como

$$N_{i+} = \sum_{j=1}^{C} N_{ij} - N_{+j}$$

N_{+j} representa el total de individuos en la columna j , calculado como

$$N_{+j} = \sum_{i=1}^{R} N_{ij} - n$$

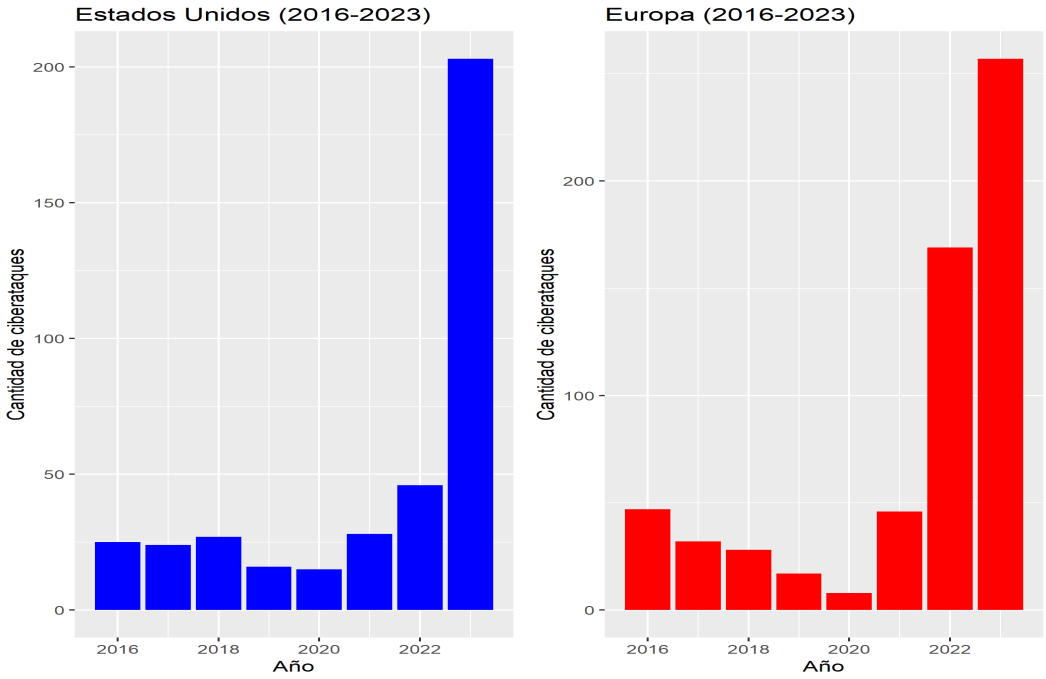
n representa el total de observaciones, calculado como

$$\sum_{i=1}^{R} \sum_{j=1}^{C} N_{ij} = n$$

RESULTADOS

En primera instancia, a partir del gráfico número 1, se puede visualizar que la distribución de datos del periodo 2016-2019 respecto a la cantidad de ciberataques nominales tanto para la región de Europa como para la de Estados Unidos no siguen una distribución normal. Además, se observa que, en el caso de Estados Unidos, las diferencias entre las cantidades nominales de ciberataques se mantienen considerablemente estables. En el gráfico de Europa, se evidencian cambios considerables, incluyendo una reducción constante en los primeros cinco años, seguida de un crecimiento exponencial. Esto puede estar relacionado con lo mencionado por Wiggen (2020), en donde el autor resalta cómo el Covid-19 generó una oleada de ciberataques en las regiones de Estados Unidos y Europa.

Figura 1: Histogramas de Ciberataques por Año

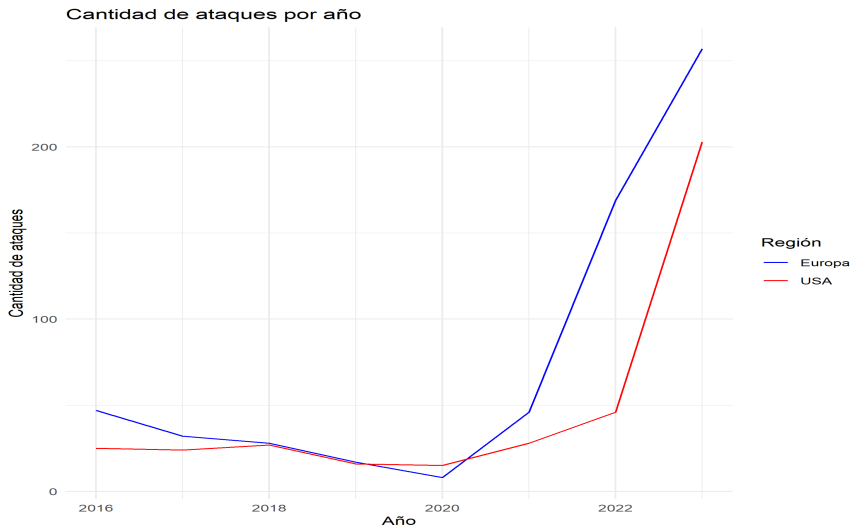


Fuente: Elaboración propia con datos del Repositorio Europeo de Ciberincidentes

Además de lo anterior, también se puede evidenciar en el gráfico 2 que, la cantidad de ciberataques presentes en Estados Unidos es notable. A pesar de que solo se trata de un país, el número de ciberataques presenta una diferencia considerablemente pequeña en comparación con toda la región de Europa, lo que refleja claramente la magnitud de los casos en Estados Unidos en comparación con otras regiones. Además, la cantidad de ciberataques en Estados Unidos tiende a ser mayor en el periodo de 2018 a 2020. Por último, se puede evidenciar cómo el año 2020

representó un cambio significativo en la cantidad de ciberataques para ambas regiones, especialmente en Europa, donde el crecimiento después de este periodo fue exponencial.

Figura 2: Gráfico de lineas de Ciberataques por Año



Fuente: Elaboración propia con datos del Repositorio Europeo de Ciberincidentes

Por otra parte, al convertir los datos a un formato de razón o proporción, permite tener valor reales de ciberataques y no solamente los valores nominales, se puede visualizar que la razón de ciberataques anual para el periodo 2016-2023 en las regiones de interés Estados Unidos y Europa, mantiene de igual manera una distribución similar a la de los datos nominales, con la ligera diferencia que los datos entre años como el caso de Estados Unidos 2016-2020 lucen aún más constantes.

Figura 3: Gráfico razón anual Estados Unidos

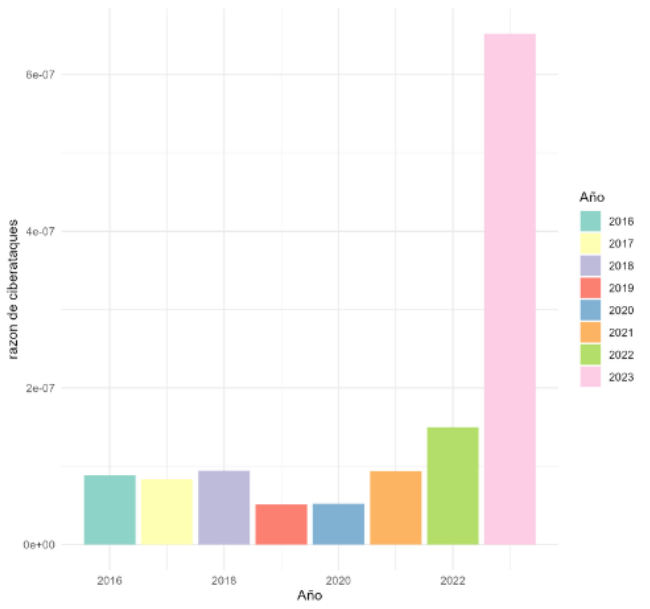
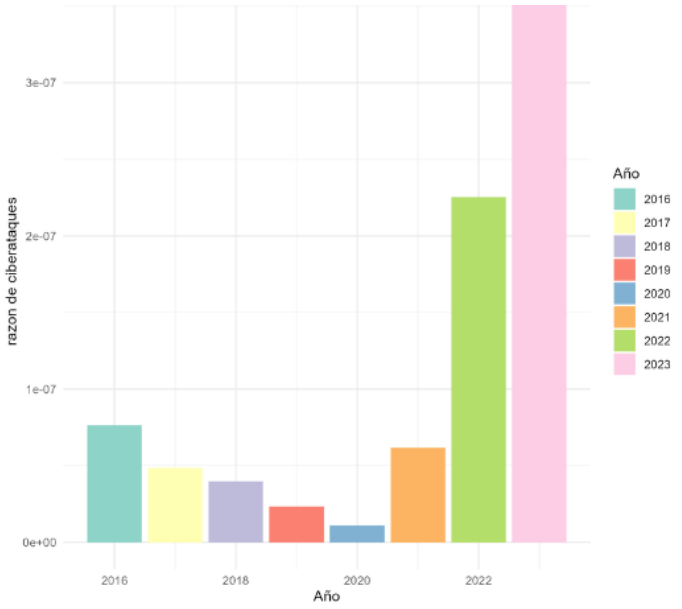


Figura 4: Gráfico razón anual Europa



Fuente: Elaboración propia con datos del Repositorio Europeo de Ciberincidentes

Luego, mediante la formación de las poblaciones (Europa y Estados Unidos) y dos grupos de fechas (antes del COVID-19 (2016-2019) y durante y después del COVID-19 (2020-2023)), y al realizar la prueba Kolmogorov-Smirnov

con respecto a la razon de ciberataques anuales entre ambos periodos para ambas regiones, se concluye lo evidenciado en **??**, la existencia de evidencia significativa para rechazar la hipótesis nula de que las distribuciones son iguales para el caso de Europa. Esto implica que las distribuciones son diferentes en los periodos de 2016-2019 y 2020-2023, lo cual indica que existe una diferencia en el periodo COVID-19 con respecto al periodo pre-COVID. Sin embargo, para el caso de Estados Unidos, no es posible rechazar la hipótesis nula, determinando que una vez que comenzó el COVID, existe un aumento significativo en la cantidad de ciberataques en la región de Europa.

Cuadro 2: Resultados de la Prueba de Kolmogorov-Smirnov en los periodos 2017-2019 y 2020-2023 para USA y Europa

Resultado de la prueba de Kologorov-Smirnovm Comparación de distribuciones cuatrimestrales Europa y USA			
Periodo 2016-2019	Periodo 2020-2023	D	p-value
Europa	Europa	0.5	0.09955
USA	USA	0.41667	0.2461

Fuente: Elaboración propia con datos del Repositorio Europeo de Ciber Incidentes.

Posteriormente, para implementar diferentes pruebas estadísticas, es importante determinar de manera confiable si las muestras para ambas regiones presentan una distribución normal. Para ello, se desarrolla una prueba de normalidad de los datos similar a la prueba Kolmogórov-Smirnov anterior, comparando cada muestra de datos del 2016-2023 para ambas regiones, con una distribución teórica normal. Se determina que las muestras para ambas regiones no siguen una distribución normal teórica, por lo que se descartan las pruebas que requieran la hipótesis de normalidad para el desarrollo de la investigación. Lo anterior se puede visualizar en los datos p obtenidos del cálculo anterior en 3

Cuadro 3: Resultados de la prueba Kolmogorov-Smirnov normalidad para Europa y USA

País	**Estadístico D**	**Valor p**
Europa	0.5	0.02259
USA	0.5	0.02259

Fuente: Elaboración propia con datos del repositorio europeo de ciberincidentes.

Considerando lo anterior, el siguiente paso es determinar la independencia de las variables. Dado que las muestras no siguen una distribución normal, se utiliza una prueba no paramétrica para evaluar la independencia. Utilizando datos de Estados Unidos y Europa durante y después del período del COVID-19 (2020-2023), así como el período previo al COVID-19 (2016-2019). Se aplica la prueba exacta de Fisher para determinar la asociación de la variable Importancia política en ambos períodos, así como para la variable nivel de impacto, que se define como medio-bajo si el índice es menor a 9 y alto si es igual o mayor a 9. Los valores p obtenidos permiten concluir que para Estados Unidos ambas variables están asociadas antes y durante/después de la pandemia. Sin embargo, para Europa, los valores p obtenidos para ambas variables son inferiores a 0.15, el nivel de significancia aceptado, por lo que no se puede concluir que las variables estan relacionadas. Los valores p específicos son: 4 y 5. Además, Hawdon y cols. (2020) aplicó la prueba chi-cuadrado para investigar relaciones entre diferentes tipos de ciberataques.

Cuadro 4: Resultados de la Prueba de Fisher de independencia en los periodos 2017-2019 y 2020-2023 para USA y Europa con respecto a la variable de importancia

Resultados de la prueba de Independencia para (2016-2019) y (2020-2023) en Europa y USA para		
Periodo 2016-2019	Periodo 2020-2023	p-value
Europa	Europa	0.0909
USA	USA	0.3427

Fuente: Elaboración propia con datos del Repositorio Europeo de Ciber Incidentes.

Cuadro 5: Resultados de la Prueba de Fisher de independencia en los periodos 2017-2019 y 2020-2023 para USA y Europa con respecto a la variable de nivel de impacto

Resultados de la prueba de Independencia para (2016-2019) y (2020-2023) en Europa y USA para		
Periodo 2016-2019	Periodo 2020-2023	p-value
Europa	Europa	0.02265
USA	USA	0.21567

Fuente: Elaboración propia con datos del Repositorio Europeo de Ciber Incidentes.

Lo anterior, lleva a la necesidad de utilizar pruebas estadísticas que no requieran de normalidad, es decir, pruebas no paramétricas que no dependan en gran medida de la independencia de los datos. Se procede a desarrollar una prueba de signos, ideal para observaciones pequeñas y no paramétricas, que no depende fuertemente de la independencia. Con lo cual, se forman dos grupos de poblaciones. Europa y Estados Unidos, y dos grupos de fechas: antes del COVID-19 (2016-2019) y durante y después del COVID-19 (2020-2023), de forma cuatrimestral. Al realizar la prueba de signos en la razón de ciberataques por usuario de internet cuatrimestral, se obtiene para la población de Europa un valor de p de 0.038 (6). Considerando una significancia del 0.15, se rechaza la hipótesis nula de igualdad de medianas y se determina la existencia de una diferencia respecto a las medianas. Para la población estadounidense, se obtiene un valor de p de 0.146(6), por lo que se puede rechazar la hipótesis nula y determinar que existe una diferencia en las medianas. Por otra parte, Hawdon y cols. (2020) usando la prueba T trató de determinar un cambio en las medias de 2 muestras; en este caso, usando la prueba de signo, se logró determinar la diferencia entre las medianas de las muestras.

Cuadro 6: Resultados de la Prueba de Signos en los periodos 2017-2019 y 2020-2023 para USA y Europa

Resultados de la prueba de signos cuatrimestral para Europa y USA para 2 muestras (2016-2019) y (2020-2023)			
Periodo 2016-2019	Periodo 2020-2023	D	p-value
Europa	Europa	3	0.03857
USA	USA	4	0.1460

Fuente: Elaboración propia con datos del Repositorio Europeo de Ciber Incidentes.

Sumado a lo anterior y aplicado al mismo grupo de población de forma cuatrimestral, se desarrolla una prueba no perimétrica para comparar las varianzas de las 2 muestras 2016-2019 y 2020-2023 para Europa y Estados Unidos 7, con lo cual por la robustez de la prueba, se decide desarrollar una prueba de Brown-Forsythe, de la cual se obtiene para el caso de Estados Unidos un p valor de 0.076, lo cual considerando el nivel de significancia establecido de 0.15, permite rechazar la hipótesis nula de igualdad en las varianzas y, por tanto, se determina que las varianzas son diferentes, de la misma manera, para la región de Europa se obtiene un p valor de 0.023, lo cual permite rechazar la hipótesis nula y, por tanto, determinar que las varianzas son diferentes.

Cuadro 7: Resultados de la Prueba de Brown-Forsythe en los periodos 2017-2019 y 2020-2023 para USA y Europa

Resultados de la Prueba de Brown-Forsythe en los periodos 2017-2019 y 2020-2023 para USA y Europa			
Periodo 2016-2019	Periodo 2020-2023	F	p-value
Europa	Europa	6.7632	0.02361
USA	USA	3.7953	0.07635

Fuente: Elaboración propia con datos del Repositorio Europeo de Ciber Incidentes.

CONCLUSIONES

El desarrollo de esta investigación se basa en la premisa de que el COVID-19 generó un impacto significativo en la sociedad, especialmente en la forma de comunicarse, relacionarse, trabajar y en el estilo de vida de las personas. Dentro de este contexto, resulta de interés explorar el ámbito de los ciberataques, un factor de suma importancia en la era tecnológica actual, con el objetivo de determinar si hubo un impacto debido a los cambios generados por la pandemia en la cantidad de ciberataques. Para esto, se seleccionaron dos regiones con alta cantidad de estos ataques: Europa y Estados Unidos.

Desarrollando la metodología elegida, se estableció una razón o proporción de los datos para comparar la cantidad de ciberataques entre los periodos seleccionados. A partir de esta razón, se realizaron varias pruebas para evaluar las distribuciones de las muestras, verificar su independencia y analizar si las distribuciones en las regiones elegidas se comportaban de manera similar. Las pruebas utilizadas incluyen Kolmogórov-Smirnov y la prueba exacta de Fisher. Los resultados iniciales indicaron que las muestras no siguen una distribución normal, se determina que existe una asociación de los datos respecto a Europa y, por último, para el caso de Europa, existe una diferencia considerable en las distribuciones.

Adicionalmente, mediante las pruebas de signos y la prueba de varianzas de Brown-Forsythe, se determinó, para un nivel de significancia del 0.15, que las medianas y las varianzas de las muestras son diferentes. Por lo tanto, se concluye que existe un impacto debido al COVID-19 en la cantidad de ciberataques en las regiones de Estados Unidos y Europa durante el periodo 2020-2023 en comparación con el periodo pre-COVID-19 (2016-2019). La diferencia en las medianas sugiere una variación en las tendencias, y la prueba de Kolmogórov-Smirnov anterior, refuerza la existencia de diferencias entre los datos, indicando la existencia de un impacto, debido al COVID-19.

La diferencia en las varianzas muestra que la dispersión de los datos respecto a la media varía entre las muestras, lo que indica un comportamiento distinto en ambos periodos para ambas regiones. Esto refuerza la idea de que el COVID-19 sí generó un impacto en cuanto a la cantidad de ciberataques.

Uno de los principales problemas relacionados con los datos es la cantidad de observaciones. Para las diferentes pruebas, se consideró una estructura cuatrimestral debido a la baja o nula cantidad de ciberataques en algunos meses. Sin embargo, este enfoque resulta en una cantidad relativamente baja de observaciones, lo que dificulta la elección de pruebas de independencia y limita el tipo de análisis que se puede realizar. Por lo tanto, se recomienda encontrar una base de datos con más observaciones, lo que permitiría el uso de un mayor número de pruebas y profundizar en las conclusiones.

Finalmente, a partir de la bibliografía consultada, no se encontraron estudios amplios para estos periodos específicos. Los estudios sobre la cantidad de ciberataques se enfocan principalmente en el primer año de la pandemia (2020), lo que no proporciona una visión general del cambio debido al COVID-19 en periodos más largos.

AGRADECIMIENTOS

Expresamos nuestro agradecimiento al profesor Maikol Solís por su constante apoyo y por proporcionar referencias cruciales para esta investigación. Su orientación y asesoramiento a lo largo de todo el proceso de desarrollo han sido fundamentales para el éxito de este trabajo. Agradecemos en especial el tiempo dedicado a resolver nuestras dudas.

Anexos

Codigo:

Se proporciona el link del repositorio donde se encuentran documentos varios entre ellos el Rscript con la explicación detallada del funcionamiento de este. `https://github.com/colomboro/Proyecto-Estadistica-2024-Grupo-9.git`

Referencias

Baz, M., Alhakami, H., Agrawal, A., Baz, A., y Khan, R. A. (2021). Impact of covid-19 pandemic: A cybersecurity perspective. *Intelligent Automation & Soft Computing*, *27*(3).

Brown MB, F. (1974). Robust tests for the equality of variances. *Journal of the American Statistical Association*, *69*(346), 364-367. Descargado de `https://www.researchgate.net/publication/24137168_Robust_tests_for_the_equality_of_variances`

Chinchilla Morales, J. (2021). Los ciberataques.

DeGroot, M., M. y Schervish. (2002). *Probability and statistics* (Vol. 2). Pearson Education, Inc.

Edition, T. (2006). Principles of epidemiology.

Eian, I. C., Yong, L. K., Li, M. Y. X., Qi, Y. H., y Fatima, Z. (2020). Cyber attacks in the era of covid-19 and possible solution domains.

EUREPOC. (01/04/2024). *EuRepoC database.* Descargado de `https://eurepoc.eu/database/` (Fecha de acceso: 01 de abril de 2024)

Fallas, J. (2012). Prueba de hipótesis. *Recuperado de: http://www. ucipfg. com/Repositorio/MGAP/MGAP*, *5*.

Hawdon, J., Parti, K., y Dearden, T. E. (2020). Cybercrime in america amid covid-19: The initial results from a natural experiment. *American Journal of Criminal Justice*, *45*(4), 546–562.

Hollander, M., Wolfe, D. A. (1999). *Nonparametric statistical methods*. Wiley-Interscience. Descargado de `https://onlinelibrary.wiley.com/doi/book/10.1002/9781119196037`

Iberdrola. (2022). *Definición ciberataque.* Descargado de `https://www.iberdrola.com/innovacion/ciberataques`

Kim, H.-Y. (2017). Statistical notes for clinical researchers: Chi-Squared test and Fisher's exact test. *Restorative dentistry & endodontics*, *42*(2), 152–155.

Kim, T. K. (2015). T test as a parametric statistic. *Korean journal of anesthesiology*, *68*(6), 540–546.

Laan, J., Junger, M., Abhishta, A., y Jonker, M. (2023). The impact of the covid-19 pandemic on phishing frequency and content. the impact of routine activities theory and a rational choice model of crime. *The Impact of Routine Activities Theory and a Rational Choice Model of Crime (January 11, 2023)*.

Lallie, H. S., Shepherd, L. A., Nurse, J. R., Erola, A., Epiphaniou, G., Maple, C., y Bellekens, X. (2021). Cyber security in the age of covid-19: A timeline and analysis of cyber-crime and cyber-attacks during the pandemic. *Computers & security*, *105*, 102248.

McHugh, M. L. (2013). The Chi-Square Test of Independence. *Biochemia medica*, *23*(2), 143–149.

Mehta CR, P. N. (1983). Exact test of significant association in contingency tables. *Comput Stat Data Anal*, *1*(2), 169-174.

Real Academia Española. (2019). *Diccionario de la lengua española.* Descargado de `https://www.rae.es/`

statista. (2024a). *Number of internet users in the united states from 2015 to 2024.* Descargado de `https://www.statista.com/statistics/265147/number-of-worldwide-internet-users-by-region/`

statista. (2024b). *Number of internet users worldwide from 2009 to 2022, by region.* Descargado de `https://www.statista.com/statistics/265147/number-of-worldwide-internet-users-by-region/`

Tawalbeh, L., Muheidat, F., Tawalbeh, M., Quwaider, M., y Saldamli, G. (2020). Predicting and preventing cyber attacks during covid-19 time using data analysis and proposed secure iot layered model. En *2020 fourth international conference on multimedia computing, networking and applications (mcna)* (p. 113-118). doi: 10.1109/MCNA50957.2020.9264301

Wiggen, J. (2020). *Impact of covid-19 on cyber crime and state-sponsored cyber activities* (Vol. 391). JSTOR.

Yadav, R., y cols. (2021). Cyber security threats during covid-19 pandemic. *International Transaction Journal of Engineering Management & Applied Sciences & Technologies*, *12*(3), 1–7.

Zibran, M. F. (2007). Chi-Squared Test of Independence. *Department of Computer Science, University of Calgary, Alberta, Canada*, 1–7.

Solicitud revisión de Proyecto de estadística para publicar

De JOSE RICARDO ACUNA GONZALEZ <JOSE.ACUNAGONZALEZ@ucr.ac.cr>
Destinatario REVISTA SERENGUETI - ESCUELA DE ESTADÍSTICA " " <REVISTASERENGUETI.EE@ucr.ac.cr>
Cc <info@adlyceum.com>, MAIKOL SOLIS CHACON <MAIKOL.SOLIS@ucr.ac.cr>, JOHAN SEBASTIAN CASTANO BUSTAMANTE <JOHAN.CASTANO@ucr.ac.cr>
Fecha 2024-06-27 08:03

Impacto del covid19 en ciberataques.pdf (~296 KB)

Estimado comité editorial:

Por este medio, les solicito respetuosamente considerar el siguiente manuscrito en la Revista Sereguenti de la Escuela de Estadística (Universidad de Costa Rica). Además, en CC se está depositando una copia a Adlyceum.

El presente estudio se titula "Impacto del COVID-19 en el número de ciberataques en las regiones de Europa y Estados Unidos en contraste con el periodo prepandémico 2016-2019". Este estudio, en particular, responde a la pregunta: ¿Cuál es el impacto de la pandemia de COVID-19 en el número de ciberataques en las regiones de Europa y Estados Unidos durante el periodo 2020-2023, en contraste con el periodo prepandémico 2016-2019? Para resolver estas preguntas se ejecutaron pruebas con respecto a la razón de ciberataques por usuario de internet para los periodos y regiones seleccionados, entre estas pruebas se encuentran: Kolmogórov-Smirnov, Prueba de signos y Brown Forsythe, las cuales muestra como se comporta la distribución, así como se comporta la media y la varianza. Además, mediante la prueba exacta de Fisher se determina la asociación entre variables.

Para este caso encontramos que existe diferencia entre el periodo prepandémico y el pospandémico, con respecto a la media y la varianza para Estados Unidos y Europa. Consideramos que este trabajo es valioso para los lectores de la Revista Serenguetti, porque permite aplicar técnicas estadísticas en un tema tecnológico y que afecta a toda la ciudadanía (dado que mucha población está intercontectada en el mundo digital). Además, es de relevancia para la realidad actual dado el auge del uso de infraestructuras digitales en todos los ámbitos de la sociedad. Para efectos administrativos, el estudiante José Acuña González es el autor para toda correspondencia. Confirmamos que este manuscrito no ha sido publicado en otro lugar y no está siendo considerado por otra revista. Todos los autores han aprobado el manuscrito y están de acuerdo con su presentación a la Revista Serenguetti.

Le saluda cordialmente

José Acuña González