UNIVERSIDAD DE COSTA RICA

FACULTAD DE CIENCIAS BÁSICAS

ESCUELA DE MATEMÁTICA

CA-303 ESTADÍSTICA ACTUARIAL I
BITÁCORAS

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

Grupo 09

Jose Ricardo Acuña González B40047

Johan Sebastián Castaño Bustamante B71744

20/06/2024

Índice

1. Bitácora 1

1.1. Pregunta de Investigación

1.1.1. Definición de la idea

Determinar el impacto del Covid-19 en el aumento de ciberataques en las regiones de Europa y Estados Unidos, en contraste con el periodo pre-pandemico 2016-2019

1.1.2. Conceptualización de la idea

Según la Real Academia Española (2019) se pueden definir los siguientes conceptos:

- Ciberataques: Conjunto de acciones dirigidas contra sistemas de información (como bases de datos o redes computacionales) con el objetivo de perjudicar a personas, instituciones o empresas. Este tipo de acción puede atentar tanto contra los equipos y sistemas que operan en la red, anulando sus servicios, como contra bases que almacenan información, siendo esta espiada, robada o, incluso, utilizada para extorsionar.
- Ciber: Indica relación con redes informáticas.
- Ataque: Acción de atacar, acometer o emprender una ofensiva.
- Aumento: Dar mayor extensión, número o materia a algo
- Impacto: Efecto de una fuerza aplicada bruscamente.
- Región : Porción de territorio determinada por caracteres étnicos o circunstancias especiales de clima, producción, topografía, administración, gobierno.

1.1.3. Identificación de tensiones

- La conclusión ofrecerá una visión general del efecto del COVID-19 en el número de ciberataques en Estados Unidos y Europa, en contraste con el periodo pre-pandemico 2017-2023, pero no necesariamente será representativa del resto del mundo.
- Implementar un método para evaluar el impacto de la pandemia del Covid-19 en la cantidad de ciberataques en las regiones de Estados Unidos y Europa, en contraste con el periodo pre-pandemico 2017-2019

- Por último, una de las principales tensiones encontradas radica en que la base de datos no contempla el valor económico que cada ciberataque generó como pérdidas, variable que en algunos de los artículos consultados se toma en cuenta para el análisis.

1.1.4. Reformulación de la idea en forma de pregunta

- ¿Cuál es el impactó del Covid-19 en el aumento de los ciberataques en las regiones de Europa y Estados Unidos, en contraste con el periodo pre-pandemico 2017-2019?
- ¿Qué método estadístico se puede emplear para evaluar el impacto del COVID-19 en el número de los ciberataques en las regiones de Estados Unidos y Europa, en contraste con el periodo pre-pandemico 2017-2019?
- ¿Cuál es la forma adecuada de visualizar el efecto del COVID-19 en el número de los ciberataques en las regiones de Europa y Estados Unidos, en contraste con el periodo pre-pandemico 2017-2019?
- ¿Resulta importante determinar el impacto del Covid-19 en el aumento de los ciberataques en las regiones de Europa y Estados Unidos, en contraste con el periodo pre-pandemico 2017-2019?

1.1.5. Argumentación de la pregunta

Pregunta: ¿Cuál es el impactó del Covid-19 en el aumento de los ciberataques en las regiones de Europa y Estados Unidos, en contraste con el periodo pre-pandemico 2017-2019?

Contraargumentos

- Lógica: Dado el tratamiento de un tema abstracto como los ciberataques, la determinación del impacto de los mismos puede ser una tarea difícil de aterrizar. Pueden existir múltiples problemas, desde la recolección de datos, la delimitación de la pregunta de investigación hasta la cuantificación del aumento.
- Ética: Pueden existir obstáculos a la hora de obtener insumos de alta confiabilidad y facilidad. Las empresas que manejan la información de ciberataques son privadas las cuales no poseen bases de datos abiertas al público.
- Emocional: AL exponer cuál es el impacto de la investigación, se puede llegar a conclusiones que no generen tanta repercusión dado que los ciberataques no son un objeto de estudio tan llamativo o disruptivo.

Argumentos

- Lógica: Existen artículos científicos los cuales respaldan el proyecto. Al utilizarlos como referencia, se brindan bases sólidas para generar análisis con la base de datos seleccionada.

- Ética: El repositorio de donde se tomó la base de datos mantiene los más altos estándares de calidad en cuanto a recolección de datos. Las Universidades de Heidelberg, Innsbruck, la Fundación de Ciencias y Política de Alemania y el Instituto de Ciber Policía de Estonia son las entidades encargadas de crear y mantener toda la información referente al repositorio. También posee la particularidad de está en constante mejoramiento, con nueva información mes a mes. Por lo que se cerciora que la información es la más actualizada dentro del campo de estudio.

- Emocional: Dada la digitalización galopante que existe en la actualidad, estos tipos de proyectos toman mayor valor para crear conciencia en la población sobre los riesgos que existen y como se deben tomar precauciones y protecciones contra los mismos. Además, la generación de conocimiento en esta área académica/profesional genera valor agregado dentro del estudio de los ciberataques.

Concluya: Basándose en artículos científicos de gran nivel y en repositorios provenientes tanto de la academia como de actores gubernamentales, se puede generar un proyecto que provea un análisis robusto el cual genere valor agregado para que se tenga mayor conciencia y se tomen precauciones ante un riesgo inminente dentro del mundo digital: los ciberataques.

Pregunta: ¿Qué método estadístico se puede emplear para evaluar el impacto del COVID-19 en el número de los ciberataques en las regiones de Estados Unidos y Europa, en contraste con el periodo pre-pandemico 2017-2019?

Contraargumentos

- Lógica: La comparación entre el periodo del Covid-19 (2020-2023) y períodos anteriores o posteriores en términos de la incidencia de ciberataques en las regiones de Europa y América del Norte presenta desafíos considerables. La pandemia ha traído consigo una disrupción sin precedentes en diversos aspectos de la sociedad, incluida la ciberseguridad, lo cual complica la identificación de un método efectivo para determinar si ha habido un impacto significativo en la cantidad de ciberataques durante este periodo con precisión.

- Ética: El enfoque seleccionado para determinar si hubo un impacto en el número de ciberataques debido al Covid-19 podría involucrar técnicas poco precisas para comparar los periodos de pande-

mia y no pandemia. Esto podría resultar en sesgos o fallos al intentar discernir si hubo un aumento o una disminución en la cantidad de ciberataques, lo que llevaría a conclusiones incorrectas.

- Emocional: Si el modelo utilizado conduce a conclusiones erróneas que no reflejan con precisión la realidad o el contexto experimentado, esto podría socavar la confianza en dicho modelo y, en consecuencia, en el análisis desarrollado.

Argumentos

- Lógica: Emplear modelos que faciliten la comparación entre dos poblaciones, como el periodo durante y fuera del Covid-19, por ejemplo, mediante el uso de pruebas de hipótesis para contrastar medias entre ambas poblaciones, proporcionaría una percepción más clara de cómo están operando estos estadísticos en ambos casos, permitiendo así determinar si ha ocurrido un cambio significativo.

- Ética: Al seleccionar y aplicar adecuadamente modelos estadísticos, como la prueba de hipótesis, se busca obtener una alta precisión en el análisis, lo que permite llegar a conclusiones más precisas y confiables sobre los fenómenos estudiados.

- Emocional: Al obtener conclusiones precisas y confiables sobre el objeto de estudio, junto con un análisis detallado, se logra reflejar con mayor fidelidad la realidad de la situación en ambas regiones, lo que contribuye a mantener la confianza en el estudio realizado.

Concluya: Es crucial que el método estadístico utilizado para evaluar el impacto del Covid-19 en el número de ciberataques en las regiones de Europa y América del Norte sea preciso y permita una comparación adecuada entre ambas poblaciones. Esto es esencial para determinar de manera confiable y representativa si ha habido un cambio significativo en la cantidad de ciberataques debido al Covid-19 en ambos territorios durante estos períodos de tiempo.

Pregunta: ¿Cuál es la forma adecuada de visualizar el efecto del COVID-19 en el número de los ciberataques en las regiones de Europa y Estados Unidos, en contraste con el periodo pre-pandemico 2017-2019?

Contraargumentos:

- Lógica: La presencia de factores externos al Covid-19 que podrían influir en el aumento del número de ciberataques en las regiones de América del Norte y Europa podría generar problemas, como un sesgo en la información, lo que dificultaría la visualización precisa del verdadero impacto del Covid-19 en los casos de ciberataques.

- Ética: Centrarse únicamente en las regiones de Europa y América del Norte puede proporcionar una visión general de la situación, pero podría no ofrecer una comprensión detallada de los lugares específicos donde se está generando o no un impacto en el número de ciberataques.
- Emocional: Examinar únicamente el número de ciberataques en las regiones de América del Norte y Europa puede impedir una comprensión completa del verdadero impacto que cada uno de estos ataques ha tenido en ambas regiones.

Argumentos

- Lógica: El análisis de los datos durante el período del Covid-19 en comparación con los datos anteriores a esta crisis y el posible cambio sustancial en el número de ciberataques permitiría visualizar y concluir que el cambio en los ciberataques está estrechamente relacionado con el surgimiento del Covid-19.
- Ética: Enfocarse exclusivamente en las regiones de Europa y América del Norte, aunque no brinda una imagen detallada de los cambios en los números de ciberataques, sí ofrece una visión general del comportamiento de estos ataques a nivel mundial, ya que estas dos regiones representan algunas de las áreas con mayor incidencia de ciberataques anuales.
- Emocional: Al observar únicamente el número de ciberataques en las regiones de América del Norte y Europa, aunque no nos proporciona una visión detallada de su impacto económico, sí nos ayuda a entender que, si existe un cambio significativo entre ambos períodos, esto implica, en términos generales, un mayor efecto. Aunque cada ataque individualmente puede tener menos impacto económico, social y político, si la diferencia entre los períodos es considerable, el conjunto de los ataques sí puede tener un mayor impacto.

Concluya: La manera adecuada de visualizar el efecto del Covid-19 en el número de ciberataques en las regiones de Europa y América del Norte debe ser una que permita análisis, visualización y comparación de los datos antes y durante la crisis del Covid-19. Esto facilitaría la determinación de la relación entre el Covid-19 y el aumento en el número de ciberataques. Además, al enfocarse en las regiones de Europa y América del Norte, se podría obtener una noción general del cambio en el número de ciberataques a nivel mundial.

Aunque enfocarse únicamente en el número de ciberataques no proporciona una visualización detallada del impacto en la sociedad, sí permite determinar si hubo un cambio en este aspecto, lo que a su vez sugiere que otros ámbitos de la sociedad también se vieron afectados.

Pregunta: ¿Resulta importante determinar el impacto del Covid-19 en el aumento de los ciberataques en las regiones de Europa y Estados Unidos en el periodo 2020-2023?

Contraargumentos

- Lógica: La relevancia del tema puede verse opacado por otros temas modernos que tomen mayor valor si se analiza el estado de la era digital donde vivimos.
- Ética: El tema expone infracciones e incumplimientos éticos desarrollados por grupos terroristas. Darles exposición puede ser una manera negativa de enfrentar el tema.
- Emocional: La importancia del análisis generado puede no generar tanto impacto para el nivel de profundidad que se quiere desarrollar.

Argumentos

- Lógica: El estudio y análisis de los ciberataques es un tema que compete a múltiples actores académicos y profesionales. Esto dado que genera muchas pérdidas económicas cuando se realiza de manera efectiva. En Costa Rica los ataques al Ministerio de Hacienda y a la Caja Costarricense de Seguro Social representaron un problema de gran magnitud para el país.
- Ética: Como estudiantes es trascendental poseer un sentido de ética con los proyectos que se desarrollan. La generación de conocimiento debe de seguir un camino que mejore el bienestar de las demás personas. Con la importancia de determinar el impacto del covid-19 en el aumento de los ciberataques se genera un análisis valioso para que se promueva conciencia en la población sobre los riesgos que existen y como se deben tomar precauciones y protecciones contra los mismos.
- Emocional: La relevancia del estudio de los ciberataques corresponde a que genera valor agregado dentro del área académica. La finalidad es generar conciencia para que se tomen precauciones dentro del mundo digital.

Concluya: La importancia del análisis del aumento de los ciberataques radica en que desprende una responsabilidad ética que se debe de mantener en el mundo digital. Esto se visualiza dado que se debe de actuar con responsabilidad, tanto en el mundo físico como en el mundo digital.

1.1.6. Argumentación a través de datos

- Fuente de información: Repositorio Europeo de Ciber incidentes (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 los mismos.
- 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 2016-2023.
- 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:

    1. ID: Variable categórica la cual identifica cada ciberataque con una identificación única.
    2. name: Variable categórica. Describe muy brevemente el ciberataque de la entrada correspondiente.
    3. Description: Variable categórica la cual describe, de manera específica, el tipo de ciberataque realizado. Cada entrada es única por el nivel de especificidad de cada ciberataque.

4. Start date: Variable categórica la cual indica la fecha en que se realizó el ciberataque. Las fechas van desde 01/01/2000 hasta 01/04/2024.
5. Accident type: Variable categórica. Indica el tipo de accidente entre Robo de información, Ransomwere, secuestro con mal uso, secuestro sin mal uso y disrupción.
6. Inclusion criteria: Variable categórica que determina el tipo de ciberataque según la infraestructura que se atacó. Los posibles valores son : infraestructura crítica u objetivo de ataque político.
7. Reciever country: Variable categórica. Indica el país que recibió el ciberataque. Se registran los 193 países inscritos en la ONU.
8. Reciever Region: Variable categórica. Indica la región en donde se ubica el país que recibió el ataque. Contempla las regiones tradicionales del mundo: 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.
9. Receiver category: Variable categórica. Indica la categoría del ciberataque entre institución estatal, corporación privada, medios de comunicación, grupos sociales, centros de investigación y otros.
10. Attributing country: Variable categórica. Indica el país que realizó el ciberataque. Contempla 60 países de múltiples regiones del mundo.
11. Attributing actor: Variable categórica. Indica el grupo terrorista que realizó el ciberataque. Contempla más de 200 grupos de ciber terroristas.
12. Unweighted cyber intensity: Variable cuantitativa. Describe, en un rango del 0 al 12, el nivel de intensidad del ciberataque realizado.
13. Economic impact exact value: Variable cuantitativa. Especifica el impacto económico, en euros, del ciberataque realizado. Como inconveniente se tiene que solo posee entradas en menos de 50 observaciones. La gran mayoría de las observaciones no tiene asignado un valor.

1.2. Revisión bibliográfica

1.2.1. Construcción de fichas de literatura

Cuadro 1: Ficha de literatura 1

Encabezado	**Contenido**
Título:	Cyber security in the age of COVID-19: A timeline and analysis of cyber-crime and cyber-attacks during the pandemic.
Autor(es):	Lallie, Harjinder Singh y Shepherd, Lynsay A y Nurse, Jason RC y Erola, Arnau y Epiphaniou, Gregory y Maple, Carsten y Bellekens, Xavier
Año:	2021
Nombre del tema:	Cibercrimen y Ciberataques durante la pandemia
Cronológica:	Marzo del 2020
Metodológica:	Descripción detallada del impacto del Covid-19 en el aumento de los ciberataques, así como formas únicas de ciberataque producto a la realidad del Covid-19.
Temática:	Análisis descriptivo
Teórica:	Implicaciones de la realidad del Covid-19 con respecto a los ciber ataques y a la ciberseguridad.
Resumen en una oración:	Los cambios generados por el Covid-19 que impulsan el número y desarrollo de ciberataques.
Argumento central:	El Covid-19 y los cambios que este genera con respecto a los ciberataques, amenazas cibernéticas, métodos nuevos de ataques cibernéticos y mejora en la ciberseguridad.
Problemas con el argumento o el tema:	Uno de los problemas encontrados por el autor es lo relacionado a diferenciar que aspecto de los ciberataques están relacionados con el Covid-19. Además de algunos problemas con la creación de la línea de tiempo planteada por el autor. Además de que los datos usados por el autor corresponden al periodo de marzo a mayo, con lo cual no se logra analizar todo el periodo del Covid-19 sino solamente esas fechas.
Resumen en un párrafo:	Se entiende en un inicio que el Covid-19 genero un impacto en el número de ciberataques en el periodo estudiado marzo-mayo del 2020. Además, se desarrolla una línea del tiempo de los diferentes ataques cibernéticos en el periodo antes mencionado. Se separa por tipo de ataque y país de origen centrándose en países como Estados Unidos, China, España, Reino Unido, entre otros que poseen considerables números de ciberataques. Se plantea que el cambio de realidad producto directo del Covid-19 genero un aumento de ciberataques.

Cuadro 2: Ficha de literatura 2

Encabezado	**Contenido**
Título:	Impact of COVID-19 Pandemic: A Cybersecurity Perspective.
Autor(es):	Baz, Mohammed y Alhakami, Hosam y Agrawal, Alka y Baz, Abdullah y Khan, Raees Ahmad
Año:	2021
Nombre del tema:	Impacto del Covid-19 en la ciberseguridad.
Cronológica:	2020
Metodológica:	Enfoque de resolución de problemas, junto a un proceso de jerarquía analítica para la realización de una evaluación cuantitativa sumado al método AHP se busca determinar el impacto de la pandemia en sector informático.
Temática:	Análisis descriptivo y cuantitativo del impacto del Covid-19 en la ciberseguridad.
Teórica:	El impacto de la realidad del Covid-19 con respecto al desarrollo, mantenimiento y cuidado de la información digital.
Resumen en una oración:	Impacto del Covid-19 en el desarrollo, implementación y mejora de la ciberseguridad.
Argumento central:	Riesgo de la información digital y mejora de la ciberseguridad en consecuencia del Covid-19.
Problemas con el argumento o el tema:	El mayor problema que se enfrentó el autor es a la hora de analizar el impacto que tuvo el Covid-19 con respecto a la seguridad informática, en la cual tuvo qué usar el método AHP para priorizar los 7 impactos más importantes.
Resumen en un párrafo:	Se abarca los cambios que el Covid-19 genero a la vida laboral, social, económica y empresarial de gran parte de la población, lo cual llevo a una mayor cantidad de información digital, que las empresas no tenían la capacidad de proteger y por tanto una mayor vulnerabilidad y robo de información. Además, se abarca de como la pandemia fue un evento que atraso la detección y respuesta contra los ataques informáticos y por tanto el impacto que estos generaron fue mayor. por último, se enfoca en la necesidad del desarrollo de la ciberseguridad, esto debido al cada vez mayor traspaso de acciones y desarrollo en el mundo digital y a lo cual gran parte de empresas e industrias se quedan por detrás.

Cuadro 3: Ficha de literatura 3

Encabezado	**Contenido**
Título:	Cyber attacks in the era of covid-19 and possible solution domains.
Autor(es):	Eian, Isaac Chin y Yong, Lim Ka y Li, Majesty Yeap Xiao y Qi, Yeo Hui y Fatima, Zahra.
Año:	2020.
Nombre del tema:	Ciberataques durante el Covid-19, los tipos más comunes y posibles soluciones.
Cronológica:	2020.
Metodológica:	Descripción del cambio de realidad debido al Covid-19 y su influencia en el aumento del numero de ciberataques.
Temática:	Análisis Descriptivo.
Teórica:	Descripción de los tipos de ciberataques, su aumento debido al Covid19, y posibles soluciones.
Resumen en una oración:	Tipos frecuentes de ciberataques durante el Covid-19 y posibles soluciones.
Argumento central:	Impacto de Covid-19 en el cuidado de la información digital
Problemas con el argumento o el tema:	La dificultad de determinar los casos de ciberataques, debido a la poca preocupación o atención de las personas.
Resumen en un párrafo:	Descripción detallada de diferentes tipos de ciberataques con mayor incidencia durante el Covid-19 y su funcionamiento. Impacto del Covid-19 en la vida laboral, educativa, social y su relación con el aumento de ciberataques. Importancia del cuidado de la información digital. Posibles soluciones para afrontar diferentes tipos de ciberataques.

Cuadro 4: Ficha de literatura 4

Encabezado	**Contenido**
Título:	The impact of COVID-19 on cyber crime and state-sponsored cyber activities.
Autor(es):	Wiggen, Johannes
Año:	2020.
Nombre del tema:	Vulnerabilidades informáticas generadas por el Covid-19 y su impacto en el numero de ciberataques.
Cronológica:	2020.
Metodológica:	Descripción de las vulnerabilidades informáticas generadas por el Covid-19 y su impacto en el aumento de los diferentes tipos de ciberataques como lo son Spear phishing, entre otros
Temática:	Análisis Descriptivo.
Teórica:	Vulnerabilidades generadas por el Covid-19, su impacto en el numero de ciberataques.
Resumen en una oración:	Aumento de diferentes tipos de ciberataques producto del Covid-19.
Argumento central:	Explicación de las vulnerabilidades generadas por el Covid-19 y el uso de estas para el uso de ataques cibernéticos.
Problemas con el argumento o el tema:	La falta de infraestructura para investigar y desarrollar medidas contra los ciberataques, limita las contramedidas posibles, así como determinar de manera exacta el número de ciberataques.
Resumen en un párrafo:	Descripción de las vulnerabilidades generadas por el Covid-19 con respecto a la información informática. Descripción de diferentes tipos de ciberataques más frecuentes durante el Covid-19 Sectores perjudicados por ciberataques durante el Covid-19 y algunas razones. Algunas medidas a considerar para tratar con el crimen cibernético.

Cuadro 5: Ficha de literatura 5

Encabezado	**Contenido**
Título:	Cyber security threats during covid-19 pandemic.
Autor(es):	Yadav, Rajesh y otros.
Año:	2021.
Nombre del tema:	Tipos de ciberataques y sectores mas afectados en consecuencia del Covid-19
Cronológica:	2020-2021.
Metodológica:	Descripción de tipos de ciberataques más frecuentes y sectores más afectados
Temática:	Análisis Descriptivo.
Teórica:	Sectores más afectados por ciberataques y tipos más frecuentes en consecuencia del Covid-19.
Resumen en una oración:	Sectores más afectados por los ciberataques y tipos de ciberataques más frecuentes durante el Covid-19.
Argumento central:	Los 9 tipos más frecuentes de ciberataques debido al Covid-19 y los sectores más afectados.
Problemas con el argumento o el tema:	La falta de datos precisos sobre la frecuencia o el aumento de ciertos tipos de ciberataques dificulta la evaluación de su magnitud y gravedad precisa.
Resumen en un párrafo:	Efecto del Covid-19 en el aumento de la virtualidad. Los sectores más afectados por ciberataques durante el Covid 19 y sus razones Las formas más frecuentes de ciberataques debido al Covid-19 y su funcionamiento. Visualización de las vulnerabilidades de los diferentes sectores de la sociedad.

Cuadro 6: Ficha de literatura 6

Encabezado	**Contenido**
Título:	Predicting and Preventing Cyber Attacks During COVID-19 Time Using Data Analysis and Proposed Secure IoT layered Model.
Autor(es):	Tawalbeh, Lo'ai y Muheidat, Fadi y Tawalbeh, Mais y Quwaider, Muhannad y Saldamli, Gokay.
Año:	2020.
Nombre del tema:	Impacto del Covid-19 en los ciberataques y explicación de contramedidas.
Cronológica:	2020.
Metodológica:	Descripción del cambio de realidad debido a Covid-19 en temas de ciberseguridad y métodos de respuesta.
Temática:	Análisis Descriptivo.
Teórica:	Los ciberataques y sus contramedidas producto del Covid-19
Resumen en una oración:	Efectos del Covid-19 en los ciberataques y mejoras en la ciberseguridad producto de este.
Argumento central:	Desarrollo y aplicación de medidas en la lucha contra los ciberataques producto del Covid-19.
Problemas con el argumento o el tema:	Desarrollar algoritmos de inteligencia artificial capaces de identificar patrones distintivos en diversos tipos de ciberataques con el fin de detectarlos con la mayor precisión posible.
Resumen en un párrafo:	Explicación del impacto del Covid-19 en la realidad cibernética. Explicación de las vulnerabilidades existentes y producto del Covid-19 en temas relacionados con la información digital. Contramedidas básicas para el manejo de los ciberataques durante el Covid-19. Desarrollo de métodos innovadores relacionados con la inteligencia para la protección y seguridad de la información digital.

Cuadro 7: Ficha de literatura 7

Encabezado	**Contenido**
Título:	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..
Autor(es):	Laan, Jip y Junger, Marianne y Abhishta, Abhishta y Jonker, Mattijs.
Año:	2023.
Nombre del tema:	Impacto del Covid-19 en los ciberataques y explicación de contramedidas.
Cronológica:	2020-2021.
Metodológica:	Análisis cuantitativo relacional, causal y descrip- tivo. Así como estimaciones mediante Mann-Whitney U Test para encontrar diferencias en la cantidad antes y durante la pandemia del Covid-19
Temática:	Análisis Descriptivo y Mann-Whitney U Test
Teórica:	El cambio en la rutina diaria y el aumento en el uso del internet y su relación en el cambio de la frecuencia de las 5 categorías de phishing usando Mann-Whitney U Test antes y durante la pandemia del Covid-19.
Resumen en una oración:	Uso de Mann-Whitney U Test en el número de ataques en las 5 categorías de phishing antes y durante la pandemia del Covid-19 para determinar la existencia de un cambio significativo en la frecuencia
Argumento central:	Descripción del cambio en la actividad rutinaria de las personas y su relación en el cambio de frecuencia de ataques phishing debido al Covid-19, así como uso de Mann-Whitney U Test para determinar este cambio significativo en las 5 categorías principales de phishing.
Problemas con el argumento o el tema:	El principal problema está relacionado con determinar una base de datos precisa que muestre los ataques phishing y la forma de ataque.
Resumen en un párrafo:	Descripción de los cambio en la vida rutinaria de las personas, así como del aumento del uso informático y su relación en el cambio de frecuencia de ataques phishing. Descripción de vulnerabilidades a ataques phishing debido al Covid-19 Uso de Mann-Whitney U Test en el número de ataques en las 5 categorías de phishing antes y durante la pandemia del Covid-19 para determinar la existencia de un cambio significativo en la frecuencia

Cuadro 8: Ficha de literatura 8

Encabezado	**Contenido**
Título:	Cybercrime in America amid COVID-19: The initial results from a natural experiment.
Autor(es):	Hawdon, James y Parti, Katalin y Dearden, Thomas E.
Año:	2020.
Nombre del tema:	Impacto del Covid-19 en el aumento del número de ciberataques y de las actividades cibernéticas.
Cronológica:	2020-2021.
Metodológica:	Análisis cuantitativo relacional y descriptivo. Así como el uso de pruebas T-test y X^2 en las categorías más destacadas de ciberataques para determinar un cambio significativo en el aumento de ciberataques y actividades cibernéticas en el periodo post-Covid-19 en comparación al pre-Covid-19.
Temática:	Análisis Descriptivo, T-test y prueba X^2.
Teórica:	El Impacto del Covid-19 en las actividades rutinarias y el aumento en el uso del internet y su relación en el aumento de ciberataques y actividades cibernéticas. Usando pruebas T-test y X^2 para medir el cambio del periodo pre-Covid-19 con el post-Covid-19.
Resumen en una oración:	Uso de X^2 Test y T-test para determinar la existencia en un cambio significativo en el aumento de ciberataques y el aumento de actividades cibernéticas.
Argumento central:	Descripción del cambio actividades rutinarias, como las actividades cibernéticas debido al Covid-19 y uso de las Prueba X^2 y T-test para determinar el impacto del Covid-19 en el aumento de ciberaataques y actividades cibernéticas en el periodo post-Covid-19 con respecto al pre-Covid-19.
Problemas con el argumento o el tema:	Problemas con falsos positivos y no cambios significativos en la mayoria de las pruebas X^2 realizadas.
Resumen en un párrafo:	Efecto del Covid-19 en las actividades rutinarias, como las actividades cibernéticas y su efecto en los cibercrímenes. Aspecto que los cibercriminales consideran para sus objetivos y actividades que generan mayor riesgo de recibir ciberataques. Uso de pruebas X^2 y pruebas T-test para determinar si existe un cambio significativo en el aumento ciberataques y en el aumento de las actividades ciberneticas en el periodo post-Covid-19 con respecto al periodo pre-Covid-19

Cuadro 9: Ficha de literatura 9

Encabezado	**Contenido**
Título:	Probability and Statistics.
Autor(es):	DeGroot, M. y Schervish, M.
Año:	2002.
Nombre del tema:	Impacto del Covid-19 en el aumento del número de ciberataques y de las actividades cibernéticas.
Cronológica:	2013.
Metodológica:	Análisis y desarrollo de temas sobre estadística y probabilidad
Temática:	Desarrollo de teorías estadísticas y probabilísticas
Teórica:	Desarrollo de conceptos y teorías estadísticas.
Resumen en una oración:	Desarrollo de conceptos, teoremas y teorías estadísticas, además de exposición de ejemplos relacionados a cada tema.
Argumento central:	Explicación de teorías, conceptos y aplicaciones estadísticas y probabilísticas.
Problemas con el argumento o el tema:	Se necesita una formación matemática previa para comprender los tópicos que expone el autor. Además, puede haber temas actualizados que no están incluidos en el texto.
Resumen en un párrafo:	Desarrollo de conceptos, teoremas y teorías estadísticas, además de exposición de ejemplos relacionados a cada tema con ejercicios dejados al lector. Específicamente se define la prueba de Kolmogorov-Smirnov como una prueba no paramétrica que se utiliza para determinar si un conjunto de datos sigue una distribución especificada.

Cuadro 10: Ficha de literatura 10

Encabezado	**Contenido**
Título:	T test as a parametric statistic.
Autor(es):	Tae Kyun Kim.
Año:	2015.
Nombre del tema:	T-test .
Cronológica:	2015.
Metodológica:	Descripción y explicación de la Prueba T-test.
Temática:	Descriptivo y Explicativo.
Teórica:	Prueba de T-test
Resumen en una oración:	Descripción del funcionamiento de la Prueba T-test
Argumento central:	T-test.
Problemas con el argumento o el tema:	Descripción única del desarrollo del T-test.
Resumen en un párrafo:	Se define a la prueba T-test como un tipo de prueba estadística que se emplea para comparar las medias de dos grupos diferentes. Generalmente se utiliza en situaciones en las que los participantes en un experimento se dividen en dos grupos independientes, y que cuentan ambos con distribución normal y varianzas iguales.

1.3. Teorías, Principios y Metodologías:

El Covid-19 ha transformado radicalmente la vida de las personas, evidenciando su vulnerabilidad ante crisis inesperadas. Este impacto se ha traducido en cambios significativos y diversas consecuencias en los ámbitos económico, social, laboral y sanitario, especialmente en las actividades rutinarias, las cuales son de interés para esta investigación.

En lo mencionado por Lallie y cols. (2021), se resalta cómo el Covid-19 ha alterado profundamente la vida de miles de millones de individuos, dando lugar a una nueva normalidad en términos de normas sociales, formas de vida, comunicación y trabajo. Esta transformación ha tenido un impacto considerable en el ámbito de la información digital, lo que ha provocado un aumento notable de los ciberataques como resultado de estos cambios. Este aumento, a su vez, ha impactado en la ciberseguridad, generando la necesidad de adaptarse a esta nueva realidad.

1.3.1. Impacto del Covid-19 en el número de ciberataques

En relación con los ciberataques, se puede considerar lo mencionado por Lallie y cols. (2021), el aumento repentino de personas trabajando desde casa debido al Covid-19 ha generado una preocupación significativa por la seguridad informática. La sociedad, según el autor, no estaba preparada para este incremento masivo del teletrabajo, lo que resultó en problemas para mantener la información segura y un aumento evidente de ciberataques debido a la vulnerabilidad de esta información.

En relación con lo anterior, Baz, Alhakami, Agrawal, Baz, y Khan (2021) señala cómo el Covid-19 ha provocado un cambio drástico en el estilo de vida de las personas. El autor destaca que, apenas un mes después de iniciada la pandemia, el mundo estaba más interconectado virtualmente que nunca, pero también más vulnerable. Además, resalta la utilización masiva de internet por parte de instituciones para la comunicación, dirección y transferencia de información, lo que ha exacerbado la fragilidad de la misma y ha contribuido al aumento de cibercrímenes en busca de esta información. Baz y cols. (2021) también menciona la dificultad en la implementación de sistemas de seguridad y su lentitud, lo cual ha influido considerablemente en el aumento de ciberataques debido a la facilidad de los mismos, así como el incremento de ciberataques relacionados con phishing y ransomware.

Por otro lado, Tawalbeh, Muheidat, Tawalbeh, Quwaider, y Saldamli (2020) señala cómo el uso de dispositivos personales por parte de las personas al trabajar desde casa aumenta el riesgo de ciberataques, ya que estos dispositivos pueden ser utilizados para acceder a diversas páginas web y realizar múltiples actividades que incrementan la probabilidad de interceptar malware y sufrir robo de información. Esto ha llevado a una alta vulnerabilidad en términos de seguridad y ha contribuido al aumento de los ciber-

ataques en las empresas que permiten o utilizan dispositivos personales para actividades laborales.

Además de lo mencionado anteriormente, Eian, Yong, Li, Qi, y Fatima (2020) destaca el aumento significativo, hasta cinco veces más, en la cantidad de ciberataques debido al Covid-19, especialmente a las vulnerabilidades generadas por este, como el incremento del trabajo desde el hogar y la mayor virtualidad, lo que genera mayores posibilidades de ser afectados por ataques cibernéticos. El autor también menciona el uso de malware y phishing en correos electrónicos relacionados con la pandemia, lo que ha generado temor en las personas y afectado a instituciones como el Banco Mundial y la Organización Mundial de la Salud, esta última afectada por la vulnerabilidad de un empleado que permitió el acceso al correo electrónico y facilitó el ataque cibernético. Además, los ciberataques se han categorizado en phishing, hacking, DoS, malware y fraudes financieros, siendo estos últimos los más frecuentes debido al Covid-19.

Por último, Wiggen (2020) destaca el uso de sitios web relacionados con el Covid-19 como medio de ciberataque, mediante la instalación de malware en estas páginas, lo que ha contribuido al aumento de ciberataques en Alemania. Además, se menciona el aumento de ciberataques relacionados con el espionaje mediante el uso de "spear phishingçon temas relacionados con el Covid-19. "Los sectores más afectados por estos ciberataques debido al Covid-19 son el sistema de salud, los servicios financieros y los medios de comunicación".(Yadav y cols., 2021)

1.3.2. Modelos usados para determinar el impacto del Covid-19 en el número de ciberataques

A partir del estudio realizado por Hawdon, Parti, y Dearden (2020), se observa que la autora 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 otro lado, en el estudio realizado por Laan, Junger, Abhishta, y Jonker (2023), se evidencia la aplicación del método de Mann-Whitney U Test por parte de la autora para determinar la existencia de cambios significativos en el número de ataques de phishing. En particular, Laan clasifica los ataques de phishing en cinco categorías y, mediante el método mencionado, identifica cambios significativos en los cinco tipos de ataques de phishing durante el periodo de Covid-19 en comparación con el periodo pre-Covid-19.

1.3.3. Prueba de Kolmogorov-Smirnov

Debido al punto anterior se evidencia que existen diferentes métodos utilices para poder evidenciar la existencia de un de un cambio significativo en el número de ciberataques en el periodo del Covid-19 con respecto al periodo pre-Covid-19, con lo cual se inicia con una prueba de Kolmogorov, la cual dirá si las distribuciones son iguales para ambas muestras, sumado a que ayudara a determinar si las muestras tienen una distribución normal, lo cual es de importancia para la utilización de la prueba T-test, por tanto como menciona DeGroot (2002), La prueba de Kolmogorov-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 una 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

1.3.4. Prueba T-test

Según T. K. Kim (2015) el T-test es un tipo de prueba estadística para comparar las medias de dos grupos o muestras independientes. Con lo cual defina 2 muestras X_1 y X_2 independientes entre sí y ambas con distribución normal y varianzas iguales, consiguientemente defina las hipótesis nulas y alternativas tal que:

$$H_0 : \mu_{x_1} = \mu_{x_2}$$

$$H_1 : \mu_{x_1} \neq \mu_{x_2}$$

con n_1 y n_2 elementos respectivamente, tal que $\hat{X_1}$ Y $\hat{Y_2}$ son las medias muéstrales correspondientes a cada muestra y sea S_1^2 Y S_2^2 la varianza correspondiente a la muestra 1 y 2 respectivamente y sea

$$s_p = \sqrt{\frac{(n_1-1)s_1^2 + (n_2-1)s_2^2}{n_1+n_2-2}}$$

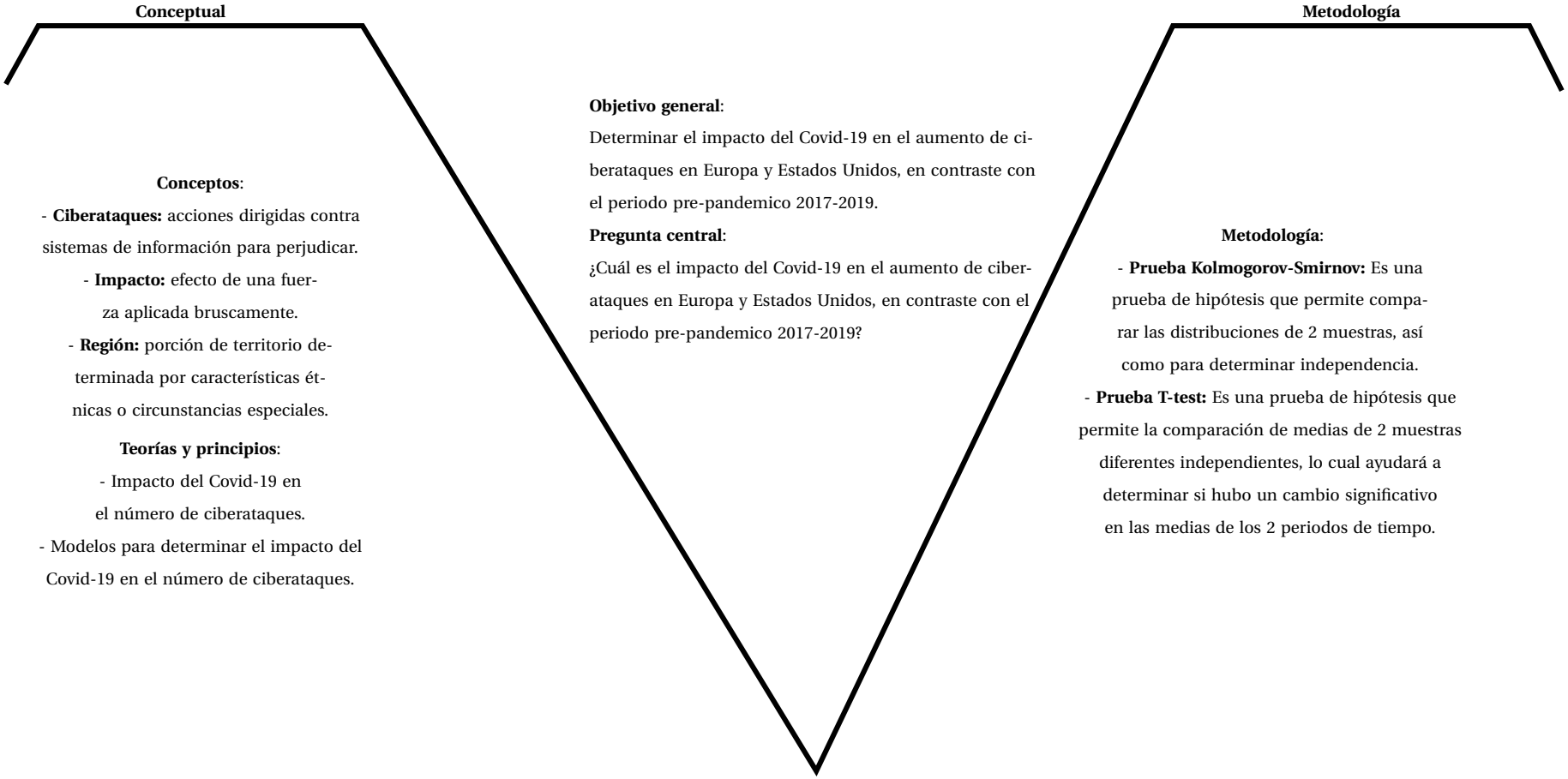
con lo cual defina el estadístico t como el estadístico de prueba, tal que:

$$t = \frac{X_1 - X_2}{sp\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}}$$

consiguientemente para poder evaluar la diferencia entre las medias, se compara el valor del estadístico con el valor teórico de la distribución t la cual consta de 4 pasos:

1. Se define el nivel de significancia.
2. Se calcula el estadístico de la prueba.
3. Se halla el valor teórico a partir de la distribución t, basado en la hipótesis nula de que las medias son iguales para ambos grupos.
4. Si el valor del estadístico de la prueba es mayor que el teórico, se descarta la hipótesis nula de la igualdad de las medias y se determina que hay una diferencia entre las medias.

1.4. Construcción de UVE de Gowin



1.5. Parte Escrita

El Covid-19 ha transformado radicalmente la vida de las personas, evidenciando su vulnerabilidad ante crisis inesperadas. En especial resalta el cambio en las actividades rutinarias, así como el aumento de la actividad digital y, por ende, la importancia y el cuidado de la información digital. Esto lleva a que la pregunta "¿Cuál es el impacto del Covid-19 en el aumento de ciberataques en las regiones de Europa y Estados Unidos, en contraste con el periodo pre-pandemico 2017-2019?"sea la opción más sólida y pertinente para el desarrollo de esta investigación, así como la más importante de determinar. Como menciona Lallie y cols. (2021), la virtualidad irrumpió ampliamente durante el 2020 y con esto los ciberataques. Por otro lado, Baz y cols. (2021) destaca la vulnerabilidad de las infraestructuras digitales frente al gran volumen repentino que significó el cambio a la virtualidad, la cual fue aprovechada por los ciber terroristas para generar ciberataques. Luego, Eian y cols. (2020) enfatiza en el cuidado que se debe fomentar a la hora de almacenar información digital y todos los protocolos de seguridad que se deben seguir para preservar, proteger y resguardar la integridad de las infraestructuras y almacenes digitales.

Además, (Wiggen, 2020) caracteriza los tipos de ciberataques más frecuentes dentro del Covid-19, entre los cuales se encuentran el robo de información, los ransomware, malware y el phishing. Por su parte, Tawalbeh y cols. (2020) señala el aumento del riesgo y número de ciberataques debido al uso de dispositivos personales en el entorno laboral remoto, lo cual multiplica las oportunidades para la infiltración de malware y el robo de información, elevando la vulnerabilidad de las empresas que lo permiten.

Wiggen (2020) también resalta cómo los espacios digitales relacionados con el Covid-19, como plataformas web infectadas de malware, contribuyen al aumento de ataques cibernéticos. El espionaje a través de "spear phishingçentrado en temas de la pandemia también ha ganado terreno, afectando principalmente a sectores críticos, como hospitales, bancos y difusores de información (Yadav et al., 2021).

Lo anterior lleva a una necesidad por encontrar métodos que permitan determinar el cambio significativo en el número de ciberataques en el periodo del Covid-19 con respecto al periodo anterior al mismo. Dentro de estos métodos se encuentran los utilizados por Hawdon y cols. (2020), donde el autor busca discernir si hay una diferencia significativa entre ambos periodos, particularmente enfocándose en el periodo posterior al Covid-19 en contraste con el anterior mediante el uso de la prueba X^2 en las diferentes categorías de ciberataques que esta analiza. Además, para evaluar posibles cambios significativos en las medias de los diversos tipos de ciberataques, el autor emplea la prueba T de Student (T-test) en ambos periodos. Además, en el estudio de (Laan y cols., 2023) se destaca el empleo del método de Mann-Whitney U Test para detectar posibles cambios importantes en la incidencia de ataques de phishing. Laan clasifica estos ataques en cinco categorías distintas y, a través del método mencionado,

identifica alteraciones significativas en los cinco tipos de ataques de phishing entre el periodo durante el Covid-19 y el periodo previo a este.

Por último, gracias a los métodos mencionados por los autores anteriores, en el desarrollo de esta investigación se decidió el uso de las pruebas T-test y Kolmogorov-Smirnov, ya que estas permiten determinar de manera precisa y clara si existe un cambio en las medias del número de ciberataques y, por tanto, permiten determinar la existencia de un cambio significativo en el periodo del Covid-19 con respecto al periodo anterior a este.

2. Bitácora 2

2.1. Parte de planificación

2.1.1. Ordenamiento de la literatura

Cuadro 11: Ordenamiento descriptivo

Organización			Literatura		
Tipo	Tema General	Tema Específico	Título	Año	Autor
Descriptivo	Cibercrimen y Ciberataques durante la pandemial	Linea de Tiempo de los ciberataques a Reino Unido durante Covid-19	Cyber security in the age of COVID-19: A timeline and analysis of cyber-crime and cyber-attacks during the pandemic	2020	Lallie, Harjinder Singh y Shepherd, Lynsay A y Nurse, Ja-son RC y Erola, Arnau y Epiphaniou, Gregory y Maple,Carsten y Bellekens, Xavier
Descriptivo	Ciberataques y contramedidas debido al Covid-19	Tipos de Vulne-rabilidades debido al Covid-19	Predicting and Preventing Cyber Attacks During COVID-19 Time Using Data Analysis and Proposed Secure IoT la-yered Model.	2020	Tawalbeh, Lo'ai y Muheidat, Fadi y Tawalbeh, Mais y Qu-waider, Muhannad y Saldamli, Gokay.
Descriptivo	Tipos de ciberataques y su aumento debido al Covid-19	Explicacion de los ciberataques más comuneras durante el Covid-19	Cyber attacks in the era of covid-19 and possible solution domains.	2020	Eian, Isaac Chin y Yong, Lim Ka y Li, Majesty Yeap Xiao y Qi, Yeo Hui y Fatima, Zahra.
Descriptivo	Vulnerabilidad debido al Covid-19	Tipos de vulne-rabilidades y su efecto	The impact of COVID-19 on cyber crime and state-sponsored cyber activities.	2020	Wiggen, Johannes.
Descriptivo	Ciberataques y sus implicaciones debido al Covid-19	10 ciberataques más frecuentes durante Covid-19	Cyber security threats during covid-19 pandemic.	2021	Yadav, Rajesh y otros
Descriptivo	Covid-19 y su impacto en ciberseguridad	7 impactos en ciberseguridad debido al Covid-19	Impact of COVID-19 Pandemic: A Cybersecurity Perspective	2021	Baz, Mohammed y otros

Cuadro 12: Ordenamiento metodológico

Organización			Literatura		
Tipo	Tema General	Tema Específico	Título	Año	Autor
Metodológico	Pruebas de Hipótesis	Kolmogorov-Smirnov	Probability and Statistics.	2002	DeGroot, M. y Schervish, M.
Metodológico	Prueba de Hipótesis	Prueba de T-test	T test as a parametric statistic.	2015	Tae Kyun Kim.
Metodológico	Descriptivo	Prueba X^2 y T-test	Cybercrime in America amid COVID-19: The initial results from a natural experiment.	2020	Hawdon, James y Parti, Katalin y Dearden, Thomas E.
Metodológico	Descriptivo	Mann-Whitney U Test	The impact of the COVID-19 pandemic on phishing fre-quency and content. The impact of Routine Activities Theory and a Rational Choice Model of crime.	2023	Laan, Jip y Junger, Marianne y Abhishta, Abhishta y Jon-ker, Mattijs.

2.2. Enlaces de la literatura

El impacto del Covid-19 ha sido significativo en todos los ámbitos de la vida, incluyendo lo económico, lo social, lo cultural y la salud. Específicamente, se han observado cambios en el estilo de vida, en el trabajo y en la forma en que las personas se comunican, entre otros aspectos importantes. Este fenómeno ha generado una nueva realidad para la población.

En relación con esto, se puede evidenciar lo mencionado en "The COVID-19 pandemic was a remarkable, unprecedented event which altered the lives of billions of citizens globally, resulting in what became commonly referred to as the 'new normal' in terms of societal norms and the way people live and work"(Lallie y cols., 2021), lo cual muestra que el Covid-19 fue tan significativo para la sociedad que incluso se habla o considera que a partir de este período hay una nueva normalidad.

El impacto del Covid-19 en diversos aspectos de la sociedad es crucial cuando se trata del aumento en el número de ciberataques. Como señala Lallie y cols. (2021), el miedo y el pánico experimentados por las personas, junto con la presión laboral y los cambios en el estilo de vida, los hacen más susceptibles y vulnerables a ataques, específicamente, en este caso, a los ciberataques.

Sumado a lo anterior y reafirmando lo mencionado anteriormente, se tiene que, "There are three factors that motivates the reason for cyber attacks; the spectacularity factor, the vulnerability factor and the fear factor"(Eian y cols., 2020), en donde se evidencia más claramente cómo la vulnerabilidad y el miedo son 2 de los factores principales de los ciberataques. Además, Lallie y cols. (2021) resalta el caso particular de desastres naturales, como el Huracán Katrina en Estados Unidos, donde después del desastre, se observó un aumento significativo en sitios web fraudulentos que buscaban realizar ataques cibernéticos con el fin de robar información o obtener beneficios económicos.

Esto sugiere la posibilidad de un aumento en el número de ciberataques considerando únicamente el impacto del Covid-19. En primera instancia, se considerará el impacto en el estilo de vida de las personas. Es particularmente importante resaltar el papel de la virtualidad y su efecto en el aumento de los ciberataques. Como se evidencia en, "Just a single month, the world became even more virtually linked and fragile than ever before. In March, the internet was suddenly used by organisations to facilitate remote communication between the large groups of home-based workplaces"(Baz y cols., 2021).

Esto evidencia el incremento de la virtualidad y cómo el mundo está más conectado que nunca, sumado a una gran vulnerabilidad debido a la poca o nula preparación del cuidado de la información por parte de la gran mayoría de la industria. Además, como se señala, el trabajo remoto desde el hogar es uno de los cambios más relevantes en relación con el trabajo. Esto se puede reflejar en

> "hackers are exploiting the vulnerabilities that appear because many people are now wor-

> king from home as well as the virus in general. The recent cyber attacks are showing up in the form of phishing emails and malware that relate to the COVID-19 virus."(Eian y cols., 2020)

en donde se muestra cómo cibercriminales se aprovechan de la poca seguridad generada por el espontáneo trabajo remoto.

Además, se refleja,

> "There are several organisations that permit their workers to use corporate applications without any tool or technique for controlling security by using their personal digital devices. Organizations' working continuity and incident response plans in perspective of industry are insufficient or even non-existent for managing a pandemic like COVID-19. An organizations' working continuity on such a scale was never expected or checked by cybersecurity officials"(Baz y cols., 2021)

En donde se refleja precisamente lo inesperado e inseguro de este cambio masivo a la virtualidad por parte del grupo de trabajo. Además, se refleja la gran poca preparación que existía en el momento para el cuidado de la información.

Sumado a lo anterior, también se puede reflejar más directamente cómo el Covid-19 influyó en el aumento de ciberataques a través de sitios web o páginas web relacionadas con temas relevantes del Covid-19. Los cibercriminales, aprovechándose del miedo generalizado debido al Covid-19 y en busca de la información más actualizada, llevaron a cabo diferentes ciberataques. Lo mencionado anteriormente puede reflejarse en,

> the COVID-19 pandemic offers an opportunity to deliberately exploit people's sense of insecurity, curiosity and their need for information for criminal or malicious activities. When someone's personal health is involved, the need for information can be easily aroused especially, when the issues at stake are protective measures, alleged treatment methods, vaccination or supposed information from government sources. In this way, internet users lose their suspicion and fall prey to scams or malware. (Wiggen, 2020)

En lo anterior se puede visualizar claramente cómo las noticias o sitios web fueron aprovechados por cibercriminales para cometer ciberataques, aprovechándose del miedo de las personas por el Covid-19.

Consecuentemente, a partir de lo anterior se puede reflejar cómo de una manera directa el Covid-19 fue aprovechado por cibercriminales para cometer ciberdelitos mediante las noticias o páginas web, así como también la importancia que tuvo la virtualidad o el crecimiento de la virtualidad en el aumento

de la vulnerabilidad y, por tanto, en el aumento de ciberataques. Sin embargo, no se ha especificado qué métodos utilizan estos cibercriminales para cometer estos delitos, con lo cual es importante comprender qué métodos fueron los más usados.

En primera instancia, se hablará del malware, del cual se menciona que,

> Malwares are continuously being spread by cyber-criminals worldwide in this present corona pandemic situation. Through many different websites and links, if the user opens the websites or clicks on a link, a trojan is being injected into the victim machine which later starts creating many different cyber issues"(Yadav y cols., 2021)

donde claramente se evidencia que el método usado como base para la mayoría de los ciberataques relacionados con páginas de noticias o páginas web fraudulentas relacionadas con el Covid-19 tiene lugar en los Malware que los cibercriminales logran introducir en los dispositivos de las víctimas para así robar información esencial.

A partir de lo anterior, se puede destacar diferentes tipos de ciberataques que están relacionados con el uso de los malware, entre los que se destaca, los ataques Phishing, los cuales,

> Phishing refers to a threat where the attackers send an email to the victim. They are made to seem authentic so that the victim follows a link provided. Following this link connects the computer to the command center, where the attackers can access any information from the victim's system. During COVID-19, people around the world have panicked. Therefore, they need any information that seems to help prevent the spread of the disease. This has allowed the attackers to use emails masked to resemble organizations such as WHO. The victims follow the links because the email seems to be authentic and beneficial. The result is an increase in successful cyberattacks. (Tawalbeh y cols., 2020)

Donde se aclara uno de los tipos más comunes de ciberataques durante la pandemia, el cual utiliza como se ha mencionado constantemente, el miedo de las personas.

Sumado a lo anterior, es importante notar que los ataques Phishing son uno de los tipos más comunes de ciberataques y por tanto es de esperar que el número de ciberataques de este tipo durante la pandemia haya crecido de manera exponencial, con lo cual se tiene,

> In the selection of the APWG data reveals a significant increase in the proportion of phishing e-mails that contained attachments. Before the pandemic, an average of 343 or 8,43% of the reported e-mails included attachments. This increase began during the pandemic, especially between May 2020 and September 2020, where an average of 1715 or 24,07% of the

> e-mails contained attachments. Using a Mann-Whitney U Test to test for difference between before and during the pandemic shows this increase is statistically significant."(Laan y cols., 2023)

Donde se evidencia de manera muy clara el crecimiento radical en el número de ciberataques relacionados con Phishing debido al Covid-19, más particularmente se puede hablar de un incremento del 15.64 % con respecto al período antes de pandemia.

Además, se menciona un período clave el cual el autor hace referencia al cual es entre mayo de 2020 y setiembre del 2020, lo cual da a entender que este período de tiempo fue en el que se visualizó un mayor impacto con respecto a los ataques Phishing. Por último, se puede evidenciar el uso de un modelo estadístico usado por el autor el cual es Mann-Whitney U test. En relación con lo mencionado anteriormente del aumento de los ciberataques, más particularmente en relación con los ataques Phishing se puede evidenciar un caso más específico de este tipo de ataque centrados en el Reino Unido, en donde se menciona, Ïndications of the extent of the UK cyber-crime incident problem experienced during the pandemic are provided by the reported level of suspect emails and fraud reported.

"By early May (07-05-20), more than 160,000 'suspect' emails had been reported to the NCSC and by the end of May (29-05-20), £4.6 m had been lost to COVID-19 related scams with around 11,206 victims of phishing and / or smishing campaigns". (Lallie y cols., 2021), en donde queda claro que para una de las potencias más importantes de Europa, el aumento de los ciberataques fue un factor determinante, así como también se presenta el impacto negativo que tuvo especialmente los ataques Phishing y el impacto de páginas web fraudulentas usadas por los cibercriminales.

Sumado al Reino Unido, también se presenta un aumento con respecto a los ciberataques producto del Covid-19 en Alemania, donde, en donde se menciona Ïn early April 2020, the Federal Office for Information Security (BSI), which is responsible for IT security in Germany, warned of an "increasing number of corona virus-related cyber attacks on businesses and citizens""(Wiggen, 2020). Lo anterior muestra cómo una institución pública de suma importancia en la seguridad alerta por el creciente número de ciberataques, no solo a ciudadanos sino incluso a compañías o negocios. Por último, incluso se presenta el caso del incremento exponencial de ciberataques en lo que respecta al gobierno estadounidense, donde se menciona,

> it has been reported that during this corona pandemic, around 600 malware attacks, 800k spam messages and 50k hits on malicious websites have been observed since May 2020 (Cook,2020). Also, starting from February to March 2020, spam emails numbers have increased 300 times and 300 % increase in malicious URLs. US is the top country for spam detection as

well as malware and the majority of target users are using them from the US"(Yadav y cols., 2021)

Donde queda claro que gran parte del mundo sufrió un incremento sustancial en la cantidad de ciberataques y más específicamente Estados Unidos presenta gran parte de los ataques relacionados a Phishing y a Malware.

Por último, es importante resaltar los métodos utilizados para determinar el impacto del Covid-19 en los ciberataques. Como menciona Laan y cols. (2023), el autor buscó determinar la diferencia en los casos de ciberataques, especialmente en los ataques de phishing. Los datos utilizados en este estudio corresponden a 6 meses antes y 4 meses después del inicio del Covid-19. El autor desarrolló una prueba de Mann-Whitney U Test para determinar si hubo un cambio en la proporción de mensajes de correos electrónicos con ataques de phishing. Además, se utilizó una prueba de T-test para dos muestras para comparar las medias entre los 6 meses anteriores a la pandemia y los 4 meses durante la pandemia. A partir de esto, el autor concluyó una afirmación sobre la proporción de ataques de phishing presentes en los correos electrónicos en ambos períodos de tiempo, centrándose en el mes de mayo.

Lo anterior es de importancia porque permite determinar una forma de calcular el posible impacto del Covid-19 con respecto a los ciberataques. En primera instancia, proporciona una idea más clara de cómo medir el impacto y determinar que lo esperado es, por lo menos con respecto a los ataques de phishing. Contrastando con lo mencionado anteriormente, se tiene lo mencionado por Hawdon y cols. (2020). En este estudio, el autor desarrolló un método de caracterización de diferentes tipos de ciberataques para 2 periodos de tiempo: durante la pandemia y antes de la pandemia. Esto se realizó mediante la aplicación de la prueba de chi-cuadrado (χ^2) para poder caracterizar, sumado a una prueba de T-test para dos períodos. Esto genera una idea firme de la implementación de la prueba de T-test en 2 periodos de tiempo para ayudar a determinar la diferencia entre el número de casos de ciberataques.

En conclusión, la pandemia de Covid-19 ha causado un aumento significativo en los ciberataques debido al miedo y la incertidumbre, así como al aumento del trabajo remoto y la dependencia de la tecnología. Los métodos estadísticos, como la prueba T-test, son importantes para comprender y abordar el cambio de estos ciberataques.

2.3. Análisis estadísticos

2.3.1. Base de datos en formato Tidy

Usando como base el capítulo 12 de R for Data Science, se adjunta la tabla que permite verificar que los datos estan en formato tidy.

ID	name	start_date
3279.00	Unknown threat actor targeted Belgian pharmaceutical chain Goed on 18 March 2022	2022-03-18
3278.00	Unknown threat actors breached US lender Nations Direct Mortgage on 30 December 2023	2023-12-30
3277.00	Unknown hackers hijacked Instagram profile of Italian Prime Minister Giorgia Meloni	2023-03-17
3270.00	ShinyHunters obtained AT&T customer data in 2021 leaking over 70 million records on 17 March 2021	2021-01-01
3273.00	Unknown Threat Actor Hit Scottish NHS Dumfries & Galloway With Cyber Attack In March 2024	2022-03-01
3272.00	Unknown actors targeted Scranton School District in Pennsylvania with ransomware on 14 March 2020	2020-03-14

inclusion_criteria	incident_type	receiver_name
Attack on critical infrastructure target(s)	Disruption; Hijacking with Misuse	Goed
Attack on critical infrastructure target(s)	Data theft; Hijacking with Misuse	Nations Direct Mortgage
Attack on (inter alia) political target(s), not politicized	Hijacking with Misuse	Giorgia Meloni
Attack on critical infrastructure target(s)	Data theft & Doxing; Hijacking with Misuse	AT&T
Attack on critical infrastructure target(s)	Data theft; Hijacking with Misuse	NHS Dumfries and Galloway
Attack on (inter alia) political target(s), not politicized	Disruption; Hijacking with Misuse; Ransomware	Scranton School District

receiver_country	receiver_region	receiver_category	receiver_category_subcode
Belgium	EUROPE; EU(MS); NATO; WESTEU	Critical infrastructure	Health
United States	NATO; NORTHAM	Critical infrastructure	Finance
Italy	EUROPE; NATO; EU(MS)	State institutions / political system	Government / ministries
United States	NATO; NORTHAM	Critical infrastructure	Telecommunications
United Kingdom	EUROPE; NATO; NORTHEU	Critical infrastructure	Health
United States	NATO; NORTHAM	State institutions / political system; Education	Civil service / administration;

initiator_country	attributing_country	unweighted_cyber_intensity	target_multiplier	weighted_cyber_intensity
Not available	Not available	0.00	Moderate - high political importance	3.00
Not available	Not available	3.00	Moderate - high political importance	3.00
Not available	Not available	2.00	Moderate - high political importance	2.00
Not available	Not available	3.00	Moderate - high political importance	3.00
Not available	Not available	3.00	Moderate - high political importance	3.00
Not available	Not available	4.00	Moderate - high political importance	4.00

Fuente: Elaboración propia con datos de Repositorio Europeo de Ciberincidentes

2.3.2. Análisis descriptivo

Para poder cumplir con los objetivos del proyecto es necesario hacer un análisis completo de los datos. Primero se muestran las siguientes tablas de frecuencia, con el fin de tener una mayor claridad de la población de estudio por condición y según el factor que se encuentre analizando dentro de la tabla. Dentro de la investigación es clave desarrollar un análisis exploratorio de los datos, Esto para comprender la información que se tiene como insumo. Primeramente, se valoran las siguientes tablas con estadísticos importantes sobre los ciberataques.

Cuadro 13: Medidas centralizadas y de dispersión

Medida	Valor
Total de Ciberataques	1300
Intensidad media	2.1
Nivel de intensidad máxima	5
Nivel de intensidad mínima	1.19
Grupos de ciberterroristas	525

Fuente: Elaboración propia con datos de Repositorio Europeo de Ciberincidentes.

Tabla de frecuencia según tipo de Ciberataque

Tipo	Intensidad	Número de ciberataques
Ransomware	4	137
Secuestro con uso indebido	2.1	390
Robo de datos	2.20	303
Disrupción	2.96	270
Robo de datos y Voxing	2.73	90
Secuestro sin uso indebido	1.19	110

Fuente: Elaboración propia con datos de Repositorio Europeo de Ciberincidentes.

Cuadro 14: Tabla de porcentajes según sector que recibió el ataque

Sector	Porcentaje del total de ciberataques
Institución del estado/sistemas políticos	49.65 %
Infraestructura Crítica	44.64 %
Otros	5.71 %

Fuente: Elaboración propia con datos de Repositorio Europeo de Ciberincidentes.

Cabe destacar que la infraestructura crítica corresponde a toda la infraestructura que brinda soporte a la industria que involucran las cuatro necesidades básicas del ser humano (alimentación, vivienda, seguridad y atención médica). El análisis de las primeras dos tablas se realizará de manera complementaria junto con las gráficas que se presentan en las siguientes páginas del análisis exploratorio de datos. El gráfico 1 muestra una clasificación con respecto al número de ciberataques realizados por país.

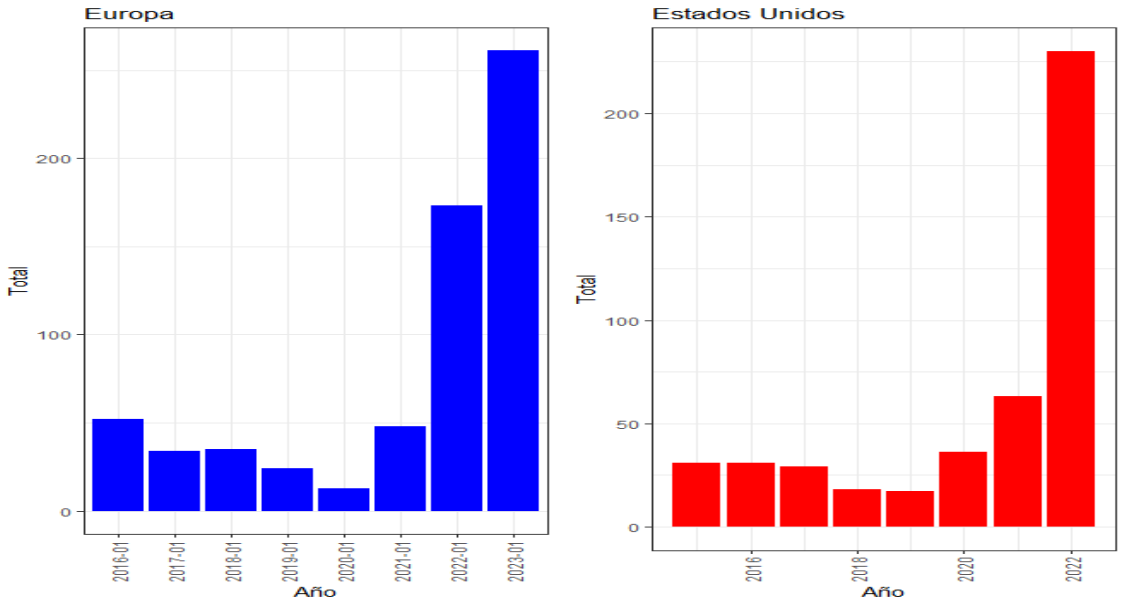
Gráfico 1: Paises de Europa por cantidad de ciberataques recibidos

Fuente: Elaboración propia con datos del Repositorio Europeo de Ciberincidentes

Los blancos más frecuentes (en orden de mayor a menor) son Rusia, Ucrania, Reino Unido y Alemania. Para el caso de los dos últimos, se valoran como países donde existen grandes centros de empresas e instituciones de gran importancia a nivel mundial; esto se traduce como un objetivo muy atractivo para los ciberdelincuentes dado que tales ciberataques representarían sacar un mayor monto económico. Las bases de datos e infraestructura digital, tanto de Alemania como de Reino Unido, se tasan en grandes sumas económicas. Por otro lado, Rusia representa un punto importante para el mundo digital. Las grandes organizaciones de ciberdelincuentes se encuentran en tal país. Así como realizan ciberataques a otros países, también están sumamente expuestos a recibirlos. Para el caso de Ucrania se valora el conflicto Ruso-Ucraniano el cual tuvo su estallido en Enero del 2022. Esto pone en evidencia la gran importancia del mundo digital. Los conflictos bélicos (así como los conocemos históricamente donde los ataques son a infraestructuras físicas y a seres humanos) se transforman a nuevos espacios donde anteriormente no se desarrollan: el mundo digital donde se atacan bases de datos, información sensible, sistemas informáticos los cuales son clave para el funcionamiento de las sociedades en el siglo XXI.

También, se puede visualizar la evolución de la cantidad de los ciberataques. A continuación, se presenta la evolución de los ciberataques realizados al continente Europeo y a Estados Unidos.

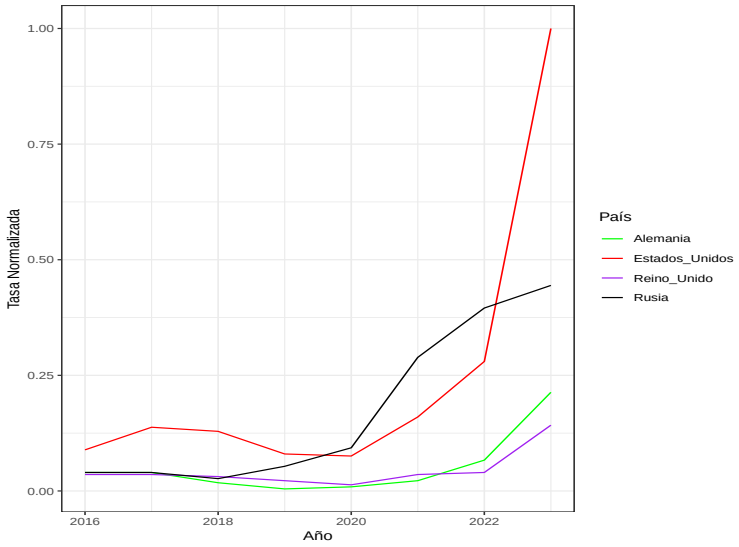
Gráfico 2: Histograma de Ciberataques por Año



Fuente: Elaboración propia con datos del Repositorio Europeo de Ciberincidentes

Conjuntamente, se puede enriquecer el análisis con el gráfico 11 al tomar los totales de ciberataques según aquellos países con mayor cantidad de ciberataques recibidos y normalizarlos en una tasa de 0 a 1, donde se toman los ciberataques por país y año y se dividen por la mayor cantidad de ciberataques recibidos por un país en un año (en este caso Estados Unidos en el 2022). De esta manera se pueden comparar en un mismo gráfico. Cabe destacar que Estados Unidos es el país que recibió mayor cantidad de ciberataques. Rusia también presenta una tasa más alta que los demás países para el periodo entre 2020 y 2022; sin embargo, es notoria la diferencia para el año 2022 donde Estados Unidos saca una gran ventaja en cuanto cantidad de ciberataques recibidos. Asimismo, se verifica como el estudio de la monotonía para los cuatro países es similar para los dos periodos de estudio dentro de la investigación (pre y post pandemia). Esto brinda una mejor óptica para poder sustentar las herramientas estadísticas que se desean ejecutar con la información recopilada del Repositorio Europeo de ciber incidentes.

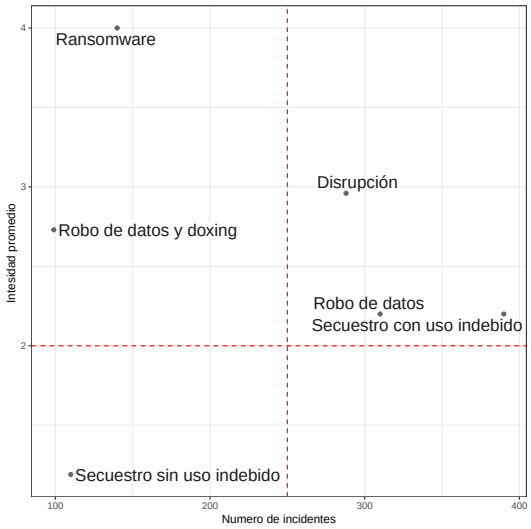
Gráfico 3: Ciberataques normalizados por país



Fuente: Elaboración propia con datos del Repositorio Europeo de Ciberincidentes

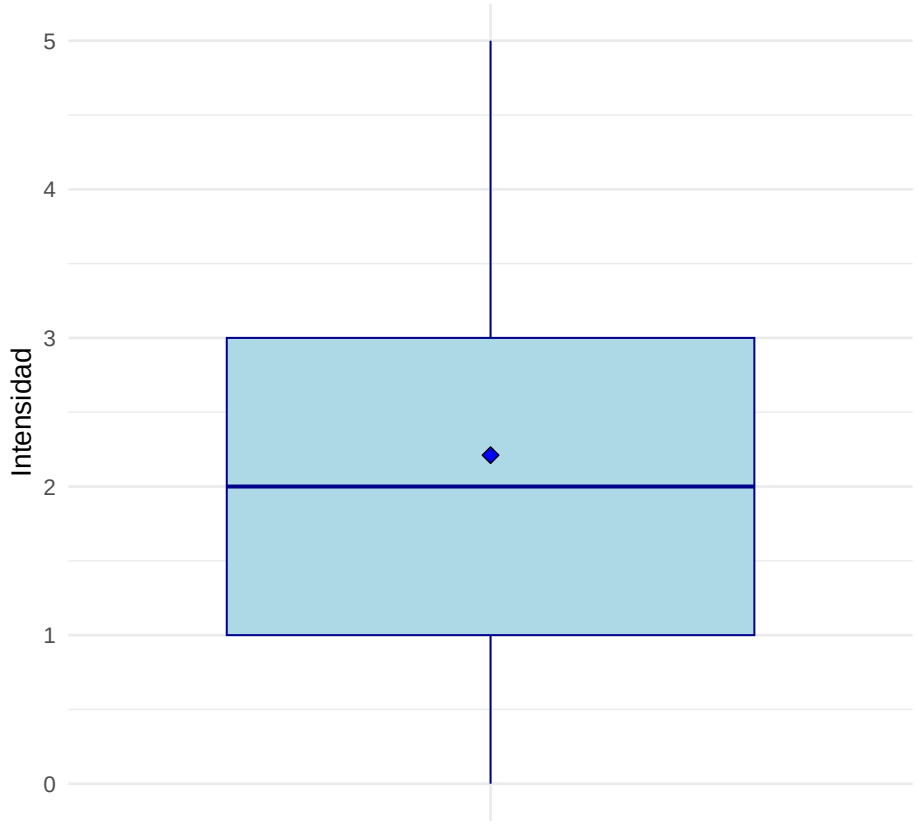
El gráfico 4 visualiza los seis tipos de ciberataques dentro de la base de datos. Se pueden reconocer tres grupos divididos por el interlineado rojo el cual es generado por las alertas establecidas por la Ciber policía de Estonia: Con una intensidad mayor a 2 se presenta un ciberataque con peligro de generar pérdidas muy elevadas. Además, recibir más de 250 ciberataques representa una vulnerabilidad dentro de los sistemas e infraestructuras digitales de las organizaciones. También, se destaca su relevancia dado que de esta manera se comprende cuales tipos de ciberataques son más comunes y cuales menos frecuentes. Con este análisis se pueden destinar recursos y esfuerzos a aquellos tipos de ciberataques que son más numerosos y además poseen una capacidad de daño mucho mayor.

Gráfico 4: Intensidad y frecuencia de los ciberataques



Fuente: Elaboración propia con datos del Repositorio Europeo de Ciberincidentes

Gráfico 5: Box plot de intensidad de ciberataques



Fuente: Elaboración propia con datos del Repositorio Europeo de Ciberincidentes

2.3.3. Propuesta metodológica

Razón de ciberataques por usuario de internet(Anual y cuatrimestral)

En primera instancia, se notó que los datos proporcionados por la base de datos del repositorio europeo de ciberincidentes están en valores nominales. En otras palabras, para poder comparar estos datos, es necesario convertirlos a 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

Cuadro 15: Número de usuarios de internet Europa (en millones) por año

Año	Usuarios de internet (millones)
2016	614.98
2017	659.63
2018	704.83
2019	727.56
2020	727.85
2021	743.60
2022	750.04
2023	726.02

Fuente: .

Cuadro 16: Número de usuarios de internet USA (en millones) por año

Año	Usuarios de internet (millones)
2016	282.10
2017	286.90
2018	286.90
2019	312.30
2020	288.10
2021	298.80
2022	307.20
2023	311.30

Fuente: statista

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.

Un ejemplo de este procedimiento se puede visualizar en:

Cuadro 17: Datos cuatrimestrales de ciberataques para USA (2016-2023)

Año	Cuatrimestre	Cantidad de ciberataques	Usuarios de internet	Razón de ciberataques por usuario
2016	1	15	94,033,333.33	1.59518E-07
2016	2	5	94,033,333.33	5.31726E-08
2016	3	5	94,033,333.33	5.31726E-08
2017	1	11	95,633,333.33	1.15023E-07
2017	2	7	95,633,333.33	7.31962E-08
2017	3	6	95,633,333.33	6.27396E-08
2018	1	12	95,633,333.33	1.25479E-07
2018	2	5	95,633,333.33	5.2283E-08
2018	3	10	95,633,333.33	1.04566E-07
2019	1	6	104,100,000	5.76369E-08
2019	2	8	104,100,000	7.68492E-08
2019	3	2	104,100,000	1.92123E-08
2020	1	9	96,033,333.33	9.37175E-08
2020	2	5	96,033,333.33	5.20653E-08
2020	3	1	96,033,333.33	1.04131E-08
2021	1	12	99,600,000	1.20482E-07
2021	2	9	99,600,000	9.03614E-08
2021	3	7	99,600,000	7.02811E-08
2022	1	10	102,400,000	9.76563E-08
2022	2	13	102,400,000	1.26953E-07
2022	3	23	102,400,000	2.24609E-07
2023	1	98	103,766,666.7	9.44427E-07
2023	2	53	103,766,666.7	5.10761E-07
2023	3	52	103,766,666.7	5.01124E-07

Fuente: Elaboración propia con datos de statista y repositorio europeo de ciberincidentes.

Prueba de Kolmogorov-Smirnov

Según DeGroot (2002), La prueba de Kolmogorov-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 una 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

A partir de lo mencionado anteriormente, se procede a implementar la prueba de Kolmogorov para determinar la relación entre las distribuciones de las muestras de datos correspondientes al periodo 2017-2019 y al periodo 2020-2023 para las regiones de Europa y Estados Unidos de forma cuatrimestral con respecto a la razón de ciberataque por usuario, respectivamente. Esto se realiza mediante la función ks.test de R, la cual requiere dos vectores numéricos que contienen los datos a comparar. La función calcula la distancia de Kolmogorov-Smirnov, previamente mencionada como D_{nm}, para realizar la comparación entre las funciones de distribución de ambos conjuntos de datos. Este valor indica la distancia entre las distribuciones, donde valores cercanos a 0 sugieren una mayor similitud entre ellas. A continuación, se presenta la implementación del código correspondiente:

```
#Prueba de Kolmogorov para comparar las distribuciones de las muestras de ambos
# periodos de tiempo, en este caso particular considerando la razon para europa
    en ambos
#periodo y luego ver la estadounidence

#caso Europa
Kolmogorov_2muestras_Europa <-
    ks.test(datos_Europa_2016_2019_cuatrimestral$razon_ciberataques_usuarios
_cuatrimestral,
    datos_Europa_2020_2023_cuatrimestral$razon_ciberataques_usuarios_
cuatrimestral)

# Imprimir el resultado
print("El resultado al comparar las distribuciones de las razones de Europa
    cuatrimestral para 2016-2019 y 2020-2023 es")
print(Kolmogorov_2muestras_Europa)

#Caso USA

Kolmogorov_2muestras_USA <-
    ks.test(datos_USA_2016_2019_cuatrimestral$razon_ciberataques_usuarios_
cuatrimestral,
    datos_USA_2020_2023_cuatrimestral$razon_ciberataques_usuarios_cuatrimestral)

# Imprimir el resultado
print("El resultado al comparar las distribuciones de las razones de USA
    cuatrimestral para 2016-2019 y 2020-2023 es")
print(Kolmogorov_2muestras_USA)
```

El cual presenta los siguientes resultados:

Cuadro 18: 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.

2.3.4. Fichas de resultados

Cuadro 19: Ficha de resultados 1

Encabezado	**Contenido**
Nombre de su hallazgo o resultado:	Determinación del tipo de ciberataque con mayor frecuencia e intensidad **??**.
Resumen en una oración:	Secuestro con uso indebido posee el mayor número de incidentes y la segunda mayor intensidad.
Principal característica:	Según toda la información recaudada se nota el patrón en los datos de que Secuestro con uso indebido es el tipo de ciberataque que combina mayor frecuencia e intensidad.
Problemas o posibles desafíos:	Se necesitaría respaldar el hallazgo del análisis exploratorio de datos con alguna prueba estadística.
Resumen en un párrafo:	Se puede analizar que Secuestro con uso indebido combina alta frecuencia y alta intensidad; sin embargo, Ransomware es el que presenta mayor intensidad de todos (Intensidad 4 en una escala del 1 al 4).

Cuadro 20: Ficha de resultados 2

Encabezado	**Contenido**
Nombre de su hallazgo o resultado:	Prueba Kolmogorov-Smirnov para determinar que las poblaciones seleccionadas no siguen una misma distribución.
Resumen en una oración:	Se desarrolla la prueba K-S y se concluye que para un nivel de significancia del 0.15 que, las muestras para el caso de Europa ni siguen una misma distribución, sin embargo para el caso de Estados Unidos, no se puede concluir que existe diferencias signiicativas con respecto a sus distribuciones.
Principal característica:	Solidifica la investigación para determinar que la distribución antes y despues del covid-19 son diferentes para el caso de Europa
Problemas o posibles desafíos:	Se necesita complementar con más análisis para poder robusteser la conlcusión que la presente investigación busca encontrar, en particular con respecto a los USA, ya que no se determino una diferencia significativa en las distribuciones
Resumen en un párrafo:	Se forman dos grupos de poblaciones (Europa y Estados Unidos) y dos grupos de fechas (antes del Covid-19(2017-2019) y durante y después del Covid-19(2020-2023)) y al realizar las prueba Kolmogorov-Smirnov se concluye que efectivamente existe evidencia significativa para rechazar la hipótesis nula para el caso de Europa, sin embargo para el caso de Estados Unidos no es posible rechazar la hipotesis nula, con lo cual, se determina que una vez que empezó el COVID, existe un aumento significativo en la cantidad de ciberataques con respecto a la region de Europa.

Cuadro 21: Ficha de resultados 3

Encabezado	**Contenido**
Nombre de su hallazgo o resultado:	Países con mayor cantidad de ciberataques recibidos.
Resumen en una oración:	Se determina que Rusia y Alemania son los paises de Europa que reciben mayor cantidad de ciberataques.
Principal característica	Se procede a establecer un orden de países que reciben ciberataques con una mayor tasa de frecuencia.
Problemas o posibles desafíos:	Se debe de ejecutar una prueba estadística para complementar el análisis realizado.
Resumen en un párrafo:	Según el gráfico **??**, se visualiza a Rusia y Alemania como los países con mayor cantidad de ciberataques dentro del tiempo de estudio seleccionado (2017-2023). Esto permite desarrollar un análisis en estos paísese para poder llegar a las causas de este fenómeno.

Cuadro 22: Ficha de resultados 4

Encabezado	**Contenido**
Nombre de su hallazgo o resultado:	Evolución de la cantidad de ciberataques a partir del 2020
Resumen en una oración:	Según **??** se tiene que apartir del 2020 aumentan los ciberataques para el total de países estudiados.
Principal característica	Se estudia la cantidad de cibertaques a nivel agregado sobre todas las observaciones contempladas
Problemas o posibles desafíos:	Como son todos los países en conjunto puede existir conclusiones distintas a la hora de desagregar la información a paises individuales.
Resumen en un párrafo:	Según gráfico **??**, se tiene que la cantidad de ciberataques ha aumentado a una tasa importante desde el 2020. Se puede distinguir el comportamiento en las época previa a la pandemia y en la época después.

Cuadro 23: Ficha de resultados 5

Encabezado	**Contenido**
Nombre de su hallazgo o resultado:	Clasificación de intensidad de ciberataques según nueva herramienta gráfica.
Resumen en una oración:	Nueva óptica de la visualización de las intensidades de los ciberataques.
Principal característica:	Permite comparar con las tablas 40 , 39 y el gráfico 4 con 5 para enriquecer el análisis.
Problemas o posibles desafíos:	Evitar Sesgos por solo utilizar información histórica. Se debe de fortalecer el análisis con múltiples herramientas estadísticas.
Resumen en un párrafo:	Mediante un gráfico de cajas, se puede generar más análisis para la variable de intensidad del ciberataque. Esta herramienta provee de información adicional, en formato de gráfico, para poder visualizar la intensidad de los ciberataques.

Cuadro 24: Ficha de resultados 6

Encabezado	**Contenido**
Nombre de su hallazgo o resultado:	Sectorización de ciberataques según tabla 14.
Resumen en una oración:	Se desarrolla una sectorización de los ciberataques según el sector que recibió el ataque.
Principal característica:	Brinda una perspectiva distinta sobre los receptores del ataque, más allá de determinar el país o región que es víctima del delito.
Problemas o posibles desafíos:	Se necesita mayor amplitud, profunidad y granularidad en los datos para poder llegar a conclusiones certeras las cuales añadan valor agregado a la investigación.
Resumen en un párrafo:	Mediante un breve análisis de datos se tiene que, dentro de los sectores de la economía, existen dos tipos que reciben el 95 % de los ciberataques: instituciones del estado e infraestructura crítica. Donde infraestructura crítica significa toda la infraestructura digital que brinda soporte a la industrias que involucran las cuatro necesidades básicas del ser humano (alimentación, vivienda, seguridad y atención médica).

2.4. Actualización de UVE de Gowin

Conceptual

Conceptos:

- **Ciberataques:** acciones dirigidas contra sistemas de información para perjudicar.
- **Impacto:** efecto de una fuerza aplicada bruscamente.
- **Región:** porción de territorio determinada por características étnicas o circunstancias especiales.

Teorías y principios:

- Impacto del Covid-19 en el número de ciberataques.
- Modelos para determinar el impacto del Covid-19 en el número de ciberataques.

Objetivo general:

Determinar el impacto del Covid-19 en el aumento de ciberataques en Europa y Estados Unidos, en contraste con el periodo pre-pandemico 2017-2023.

Pregunta central:

¿Cuál es el impacto del Covid-19 en el aumento de ciberataques en Europa y Estados Unidos, en contraste con el periodo pre-pandemico 2017-2023?

Metodología

Metodología:

- **Prueba Kolmogorov-Smirnov:** Es una prueba de hipótesis que permite comparar las distribuciones de 2 muestras, así como para determinar independencia.
- **Prueba T-test:** Es una prueba de hipótesis que permite la comparación de medias de 2 muestras diferentes independientes, lo cual ayudará a determinar si hubo un cambio significativo en las medias de los 2 periodos de tiempo.

Objetivos específicos:

- Identificar las posibles implicaciones de la crisis sanitaria del Covid-19 en el número de ciberataques en Europa y Estados Unidos durante el periodo 2020-2023, en comparación con el periodo anterior a la pandemia, 2017-2019.
- Examinar las características de los ciberataques más comunes y prominentes en Estados Unidos y Europa durante el periodo 2017-2023 con el fin de identificar las posibles causas del aumento en su frecuencia debido al impacto del Covid-19.
- Hallar un modelo estadístico que facilite la visualización y el cálculo preciso del impacto del Covid-19 en el numero de ciberataques en Europa y Estados Unidos, contrastando con el periodo pre-pandémico de 2017-2019.

3. Bitácora 3

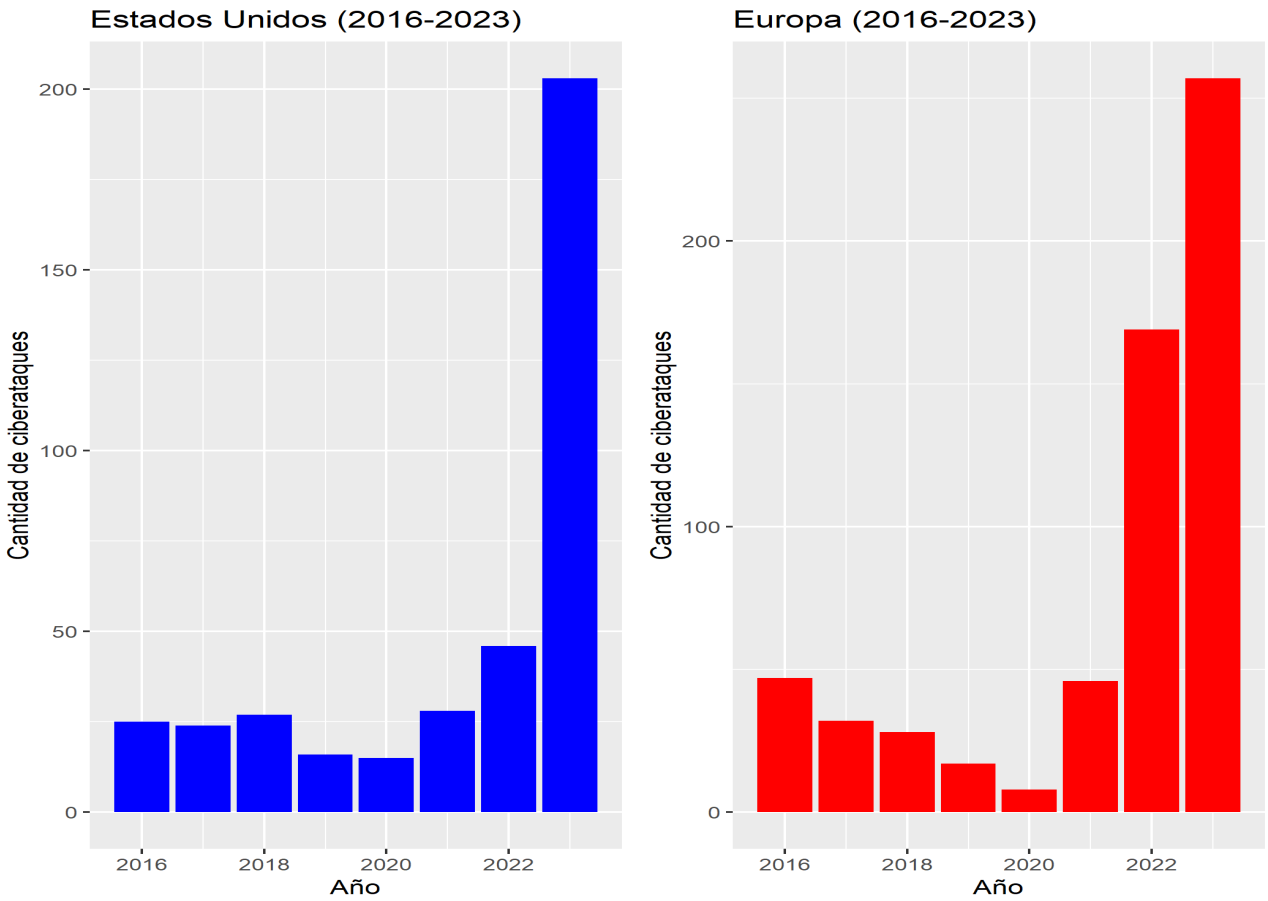
3.1. Fichas de resultados del análisis

Para responder a la pregunta de investigación, se procedió a realizar un gráfico de barras que muestra el recuento de la cantidad de ciberataques en Estados Unidos y Europa en el periodo de 2016-2023. A continuación, se presentan la ficha con los resultados respectivos y el gráfico de referencia.

Cuadro 25: Ficha de resultados 1

Encabezado	**Contenido**
Nombre de su hallazgo o resultado:	Diferencias en la cantidad de ciberataques por año
Resumen en una oración:	Se evidencian diferencia en la distribución para ambas regiones de forma relevante desde el 2021 y indicios para esperar que la ditribución no se comporta de manera normal
Principal característica:	Permite comparar las diferentes cantidades de ciberataques nominales por región.
Problemas o posibles desafíos:	Gran parte de las diferencias se concentran en los años 2022 y 2023
Resumen en un párrafo:	En primera instancia, visualmente se puede evidenciar como la distribución se espera que tanto para la región de Europa como la de Estados Unidos no sean normales, además se puede visualizar que para el caso de Estados Unidos, las diferencia entre las cantidades nominales se mantienen considerablemente estable, con respecto al gráfico Europeo se evidencian cambios considerables, además de que se evidencia una reducción constante en los primeros 5 años para luego tener un crecimiento exponencial.

Gráfico 6: Histogramas de Ciberataques por Año



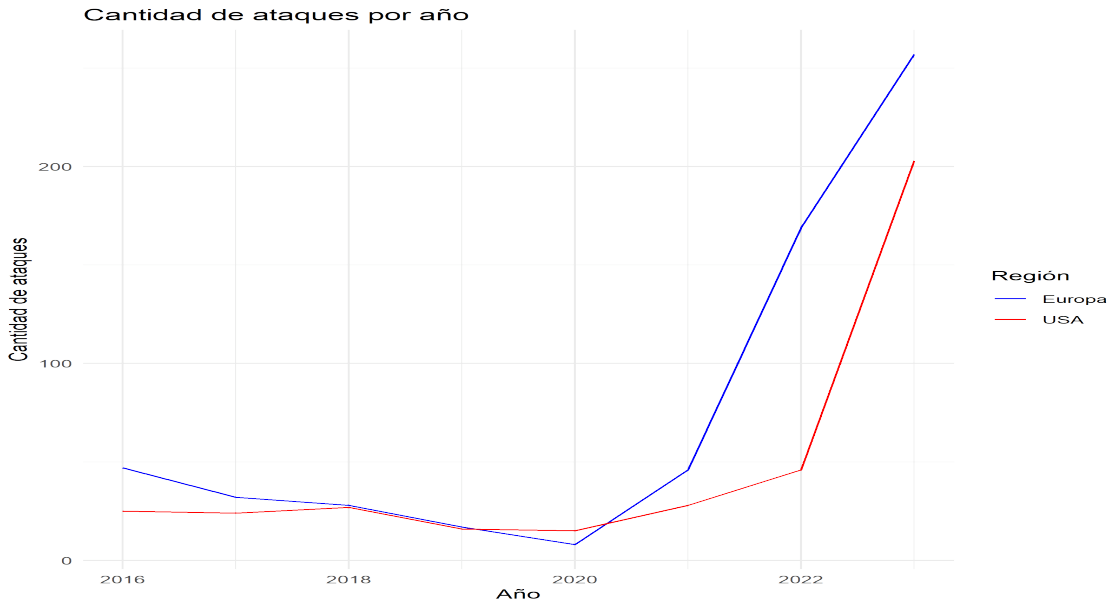
Fuente: Elaboración propia con datos del Repositorio Europeo de Ciberincidentes

A continuación se presenta una idea similar a la anterior, con la diferencia de que los siguientes gráficos y la ficha de resultados buscan realizar una comparación de la cantidad de ciberataques entre las regiones de interés, que en este caso corresponden a Estados Unidos y Europa. En primer lugar, el gráfico de líneas ofrece una visión general de los ciberataques en ambas regiones. Posteriormente, el mapa de Europa, que muestra el número de ciberataques por país, proporciona una visión más específica relacionada con la cantidad y distribución de ciberataques en Europa.

Cuadro 26: Ficha de resultados 2

Encabezado	**Contenido**
Nombre de su hallazgo o resultado:	Diferencias entre los ciberataques de USA y Europa
Resumen en una oración:	Se evidencia en primera instancia, como en términos generales la cantidad de ciberataques es mayor en Estados Unidos que en Europa
Principal característica:	Permite comparar las diferentes cantidades de ciberataques nominales entre ambas regiones de interés.
Problemas o posibles desafíos:	La separación de los datos para la región de Europa fue un reto debido a la diferencia en los nombres usados para esta región.
Resumen en un párrafo:	En primera instancia, es notable la gran cantidad de casos de ciberataques presentes en Estados Unidos. A pesar de que solo es un país, el número de ciberataques presenta una diferencia relativamente pequeña en comparación con toda la región de Europa, lo cual 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.

Gráfico 7: Gráfico de lineas Ciberataques por Año



Fuente: Elaboración propia con datos del Repositorio Europeo de Ciberincidentes

Gráfico 8: Paises de Europa por cantidad de ciberataques recibidos en el periodo 2016-2023

Fuente: Elaboración propia con datos del Repositorio Europeo de Ciberincidentes

Los siguientes gráficos y la ficha de resultados muestran la razón de ciberataques por usuario de Internet. Esto se hace con el fin de permitir una comparación precisa, evitando que un incremento en el número de usuarios de Internet sea el factor detrás del aumento en los ciberataques, y enfocándose en aspectos directamente relacionados con el COVID-19. Esta razón o proporción, vista en los gráficos, evidencia cómo, a pesar de los cambios en la cantidad nominal, la distribución de los datos se mantiene constante. Esto sugiere la existencia de una diferencia en la cantidad de ciberataques atribuible al COVID-19.

Cuadro 27: Ficha de resultados 3

Encabezado	**Contenido**
Nombre de su hallazgo o resultado:	Razón de ciberataque por usuario de internet de forma anual y cuatrimestral
Resumen en una oración:	Calculo de la razón de ciberataques por usuario de internet de forma anual y cuatrimestral para la comparación de los datos.
Principal característica:	Permite comparar las diferentes muestras por año, ya que al dividir los ciberataques entre la cantidad de usuarios de internet, da una proporción de cuantos ciberataques son por persona y por tanto permite que los datos sean comparables.
Problemas o posibles desafíos:	Dificultad de encontrar la cantidad de usuario de internet en las regiones de interés
Resumen en un párrafo:	Se calcula la razón de ciberataques anual y cuatrimestral para el periodo 2016-2023 de las regiones de interés, como la división entre la cantidad de ciberataques de forma anual y cuatrimestral y la cantidad de usuarios de internet de forma anual y cuatrimestral. De esta manera, esta razón representa la proporción de ciberataques por persona que usa internet, lo cual permite la comparación de los datos en diferentes momentos. Además, proporciona una visualización real del periodo con la mayor proporción de ciberataques.

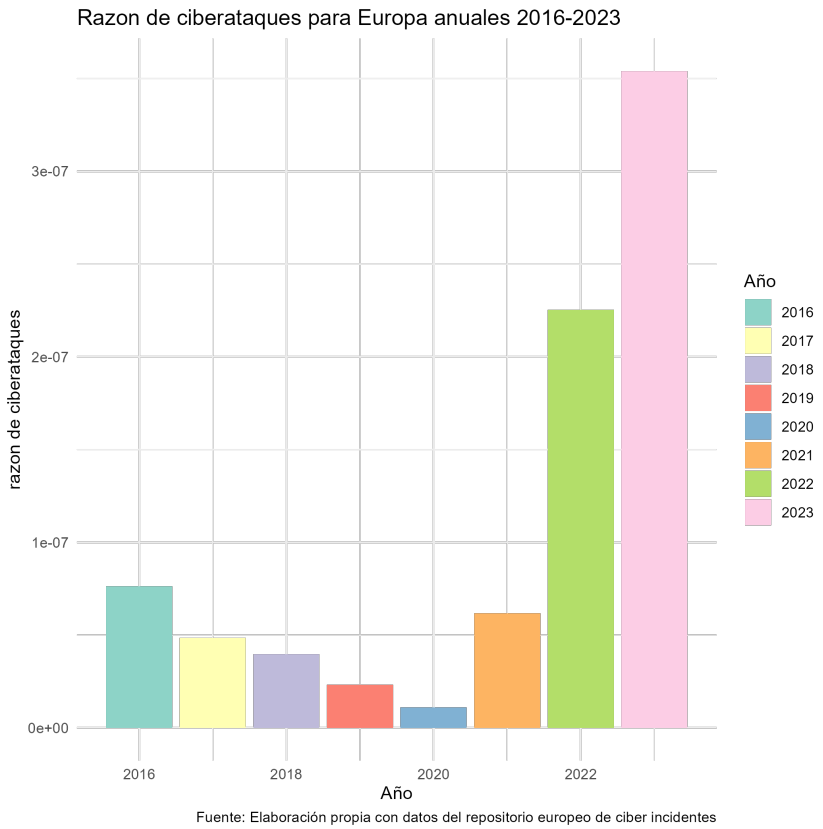


Gráfico 9

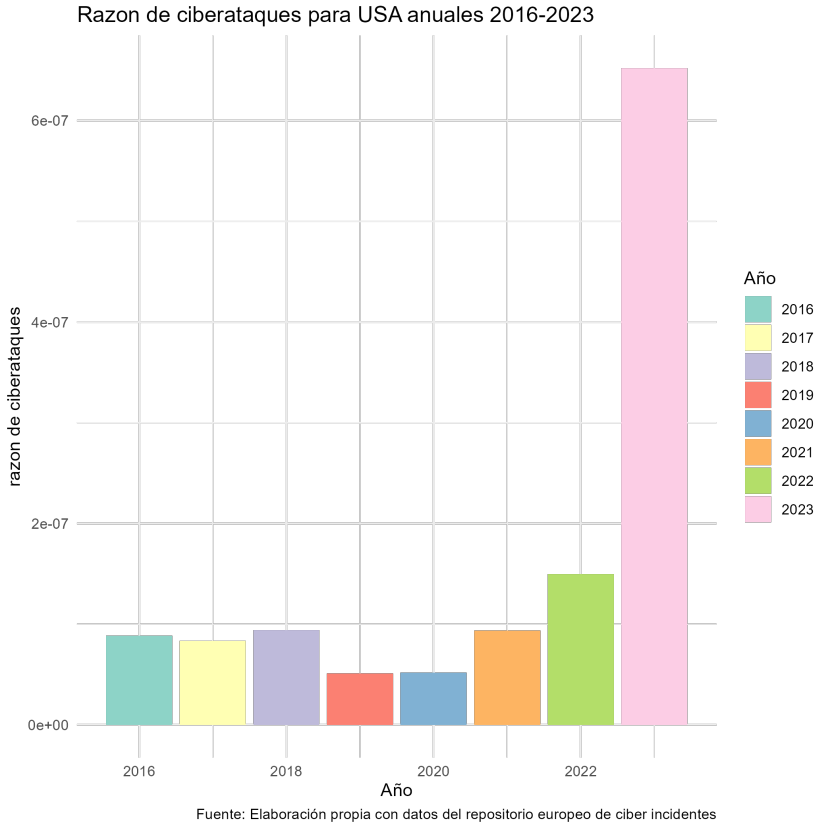


Gráfico 10

El siguiente cuadro y ficha de resultados, muestra la relación existente entre las distribuciones de los dos periodos de tiempo de interés 2016-2019 y 2020-2023 para las regiones de Estados Unidos Y Europa esto usando la prueba de Kolmogorov-Smirnov para 2 muestras.

Cuadro 28: Ficha de resultados 4

Encabezado	**Contenido**
Nombre de su hallazgo o resultado:	Prueba Kolmogorov-Smirnov para determinar que las poblaciones seleccionadas no siguen una misma distribución.
Resumen en una oración:	Se desarrolla la prueba K-S y se concluye que para un nivel de significancia del 0.15 que, las muestras para el caso de Europa no siguen una misma distribución, sin embargo, para el caso de Estados Unidos, no se puede concluir que existe diferencias significativas con respecto a sus distribuciones.
Principal característica:	Solidifica la investigación para determinar que la distribución antes y después del covid-19 son diferentes para el caso de Europa
Problemas o posibles desafíos:	Se necesita complementar con más análisis para poder robustecer la conclusión que la presente investigación busca encontrar, en particular con respecto a los USA, ya que no se determinó una diferencia significativa en las distribuciones
Resumen en un párrafo:	Se forman dos grupos de poblaciones (Europa y Estados Unidos) y dos grupos de fechas (antes del Covid-19(2017-2019) y durante y después del Covid-19(2020-2023)) y al realizar las prueba Kolmogorov-Smirnov para ambas muestra, se concluye que efectivamente existe evidencia significativa para rechazar la hipótesis nula de que ambas muestras siguen una misma distribución para el caso de Europa, sin embargo para el caso de Estados Unidos no es posible rechazar la hipótesis nula, con lo cual, se determina que una vez que empezó el COVID, existe un aumento significativo en la cantidad de ciberataques con respecto a la región de Europa.

Cuadro 29: 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.

El siguiente cuadro y ficha de resultados tienen como objetivo determinar la normalidad de los datos o muestras. Esto se debe a que la normalidad es esencial para el uso de pruebas estadísticas, por lo que es crucial establecer si los datos presentan una distribución normal. Para ello, se utilizó el método Kolmogorov-Smirnov, en el cual se estableció como distribución teórica la distribución normal, y se comparó con la distribución de la muestra anual para Europa y Estados Unidos en los periodos 2016-2023.

Cuadro 30: Ficha de resultados 5

Encabezado	**Contenido**
Nombre de su hallazgo o resultado:	Prueba de Kolmogorov-Smirnov para determinar la normalidad de las muestras.
Resumen en una oración:	Se desarrolla la prueba K-S y se determina que las muestras no siguen una distribución teórica normal
Principal característica:	Permite la justificación para el uso de pruebas no-perimétricas
Problemas o posibles desafíos:	Determinar las mejores pruebas no perimétricas para la muestra
Resumen en un párrafo:	Se forman dos grupos de poblaciones (Europa y Estados Unidos) para 2016-2023. Al realizar la prueba de Kolmogorov-Smirnov para ambas poblaciones con respecto a una distribución teórica normal para la variable razón de ciberataques por usuario de internet anual, se rechazó la hipótesis nula de que los datos siguen una distribución normal. Por lo tanto, se concluye que las distribuciones de las poblaciones no son normales.

Cuadro 31: 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.

A partir de haber determinado que los datos no siguen una distribución normal, se decidió usar pruebas no paramétricas, las cuales no requieren la normalidad como supuesto para poder ejecutarse. En este caso, se utilizó la prueba Chi-cuadrado, cuyos resultados y tabla se presentan a continuación.

Cuadro 32: Ficha de resultados 6

Encabezado	**Contenido**
Nombre de su hallazgo o resultado:	Prueba chi-cuadrado para determinar la independencia de las muestras
Resumen en una oración:	Se desarrolla la prueba chi-cuadrado y se determina que no hay evidencia significativa para rechazar la hipótesis nula y por tanto no se puede determinar que los datos son independientes.
Principal característica:	Permite determinar la necesidad del uso de pruebas que no requieran independencia o que no sean tan volubles con respecto a la independencia, como es el caso de la prueba de signos.
Problemas o posibles desafíos:	La cantidad de observaciones de la muestra puede generar problemas
Resumen en un párrafo:	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 chi-cuadrado en la razón de ciberataques por usuario de internet cuatrimestral, se obtiene para ambas poblaciones p mayores a 0.25 y por tanto no se puede rechazar la hipótesis nula de dependencia.

Cuadro 33: Resultados de la prueba Chi-cuadrado para Europa y USA cuatrimestral para 2 muestras (2016-2019) y (2020-2023)

País	**Estadístico X-squared**	**Grados de libertad (df)**	**Valor p**
Europa	96	88	0.2625
USA	120	110	0.2421

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

Por último, al haber definido que los datos no siguen una distribución normal y no poder descartar la posibilidad de que estén relacionados, se decidió utilizar una prueba no paramétrica que no se vea influenciada significativamente por la independencia de los datos y que requiere una cantidad pequeña de observaciones. En este caso, se optó por la prueba de signos, cuyos resultados y cuadro se presentan a continuación.

Cuadro 34: Ficha de resultados 7

Encabezado	**Contenido**
Nombre de su hallazgo o resultado:	Prueba de signos para determinar la diferencia en las medianas de 2 muestras.
Resumen en una oración:	Se desarrolla la prueba de signos y se determina que para un nivel de significancia del 0.15 se rechaza la hipótesis nula y se determina que si existe diferencia en las medianas.
Principal característica:	Prueba robusta no perimétrica que no depende de independencia entre las muestras.
Problemas o posibles desafíos:	considerar una significancia mayor al 0.05
Resumen en un párrafo:	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. Considerando una significancia del 0.15, se rechaza la hipótesis nula de igualdad de medianas y se determina la existencia de una diferencia con respecto a las medianas. Ahora, con respecto a la población estadounidense, se obtiene un valor de p de 0.146. Por tanto, se puede rechazar la hipótesis nula y determinar que existe una diferencia con respecto a las medianas.

Cuadro 35: 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.

Cuadro 36: Ficha de resultados 8

Encabezado	**Contenido**
Nombre de su hallazgo o resultado:	Existencia de una diferencia en las medianas de ambas regiones con respecto a las muestras 2016-2019 y 2020-2023
Resumen en una oración:	A partir de la prueba de signos, se concluye que existe una diferencia en las medianas y, por lo tanto, se infiere la existencia de una diferencia en la cantidad de ciberataques entre las regiones de Europa y Estados Unidos durante el periodo de COVID-19 (2020-2023) en comparación con el periodo pre-COVID (2016-2019)
Principal característica:	Prueba robusta no perimétrica que no depende de independencia entre las muestras.
Problemas o posibles desafíos:	considerar una significancia mayor al 0.5
Resumen en un párrafo:	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 que existe una diferencia con respecto a las medianas y por tanto se concluye la existencia de un impacto en el número de ciberataques debido al Covid-19

3.2. Ordenamiento de los elementos de reporte

Cuadro 37: Ordenamiento de elementos de reporte

Primarios	Secundarios
Teoría, aplicaciones programacionales y resultados de prueba χ^2	Teoría sobre el análisis de la ciberatques en épocas de covid (Lallie y cols., 2021)
Teoría y aplicaciones programacionales de las pruebas de signo	Teoría sobre el impacto del covid en los crimenes cibernéticos (Wiggen, 2020)
Teoría, aplicaciones programacionales y resultados de prueba Kolmogorov Smirnov para determinar que las poblaciones no siguen una misma distribución y para determinar la normalidad de las muestras	Cálculo de razon de ciberataque por usuario de internet (statista, 2024a)
Razon de ciberataques por usuario de internet(Anual y cuatrimestral)	Teoría sobre el modelo de crimen cibernético en pandemia (Laan y cols., 2023)
	Resultado de Yadav y cols. (2021) sobre las amenazas cibernéticas
	Resultados del análisis descriptivo
	Resultado del uso de prueba Chi-cuadrado y prueba T para determinar la existencia del impacto del covid-19 en el numero de ciberataques Hawdon y cols. (2020)

Cuadro 38: Ordenamiento de elementos de reporte

Sección	Temas a tratar
Introducción	1. Teoría de Lallie y cols. (2021): ciberataques en épocas de covid (Secundario). 2. Teoría de statista (2024a): crimen cibernetico (Secundario) 3. Teoría (Wiggen, 2020): impacto de covid en crímenes cibernéticos. 4. Resultado de Yadav y cols. (2021): Amenazas cibernéticas.
Metodología	1.Razon de ciberataques por usuario de internet(Anual y cuatrimestral)(Primario) 2. Prueba χ^2 (Primario). 3. Prueba de signos (Primario) 4. Prueba Kolmogorov Smirnov (Primario).
Resultados	1. Resultados del análisis descriptivo (secundario). 2. Resultados de razon de ciberataques por usuario de internet.(secundario) 3. Resultados de prueba Kolmogorov-Smirnov.(Primario) 4. Resultado de la prueba χ^2 (Primario). 5. Resultado de la Prueba de Signos (Primario). 6. Resultado del uso de prueba Chi-cuadrado y prueba T para determinar la existencia del impacto del covid-19 en el numero de ciberataques(secundario)

3.3. Parte de escritura

3.3.1. Introducción

Este 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 del Covid-19 (2020-2023) en contraste con el periodo pre-Covid (2016-2019). Para esto, es fundamental conocer los conceptos relevantes que pueden abarcar este trabajo, entre los cuales se encuentran:

Conceptos

- Ciberataque: “Es la explotación deliberada de sistemas informáticos, empresas y redes dependientes de la tecnología. Estos ataques utilizan códigos maliciosos para alterar la lógica o los datos del ordenador, lo que genera consecuencias perjudiciales que pueden comprometer información y provocar delitos cibernéticos, como el robo de identidad”.(Chinchilla Morales, 2021)

- Región: “Porción de territorio determinada por caracteres étnicos o circunstancias especiales de clima, producción, topografía, administración, gobierno”.(Real Academia Española, 2019)

- Pandemia: Según Edition (2006) una pandemia, se refiere a una epidemia que se ha extendido por varios países o continentes y que suele afectar a un gran número de personas. Además Edition (2006) también menciona como una epidemia se refiere a un aumento, a menudo repentino, en el número de casos de una enfermedad por encima de lo que normalmente se espera en la población de esa zona.

Consiguientemente, es importante resaltar que la pandemia del Covid-19 ha transformado radicalmente numerosos aspectos de la vida cotidiana y profesional, provocando una aceleración sin precedentes en la adopción de tecnologías digitales. Sin embargo, esta transición ha expuesto nuevas vulnerabilidades y ha incrementado significativamente la cantidad e intensidad de los ciberataques. Mediante un análisis de literatura científica se inicia el desarrollo el estudio del impacto del covid-19 en el aumento de los ciberataques en Estados Unidos y Europa.

Primeramente, el estudio Lallie y cols. (2021) proporciona 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.

Lallie y cols. (2021) también señalan que los ataques de phishing, ransomware y otras formas de cibercrimen se volvieron más comunes, con los atacantes adaptando sus estrategias para explotar el miedo y la incertidumbre generados por la pandemia. Las instituciones de salud y los organismos gubernamentales, en particular, fueron blanco de ataques debido a la urgencia y la sensibilidad de los datos manejados.

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.

Yadav y cols. (2021) abordan las diversas amenazas de ciberseguridad que emergieron durante la pandemia, destacando que el rápido cambio hacia el teletrabajo expuso 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 y de invertir en tecnologías que permitan la detección y mitigación de amenazas en tiempo real.

Laan y cols. (2023) ofrecen una perspectiva sobre cómo la frecuencia y el contenido de los ataques de phishing se adaptaron durante la pandemia. Utilizando la teoría de actividades rutinarias y un modelo de elección racional del crimen, los autores argumentan que los cambios en las rutinas diarias y el aumento en la actividad en línea crearon oportunidades adicionales para los ataques de phishing.

El estudio Laan y cols. (2023) muestra que los correos electrónicos de phishing se volvieron más personalizados y contextualmente relevantes, con temas relacionados con la COVID-19, como actualizaciones de salud, ayudas gubernamentales y teletrabajo. Estos correos electrónicos explotaron la necesidad urgente de información y la ansiedad generalizada, aumentando la probabilidad de éxito de los ataques.

Además, A partir del estudio realizado por Hawdon y cols. (2020), se observa que la autora 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 otro lado, en el estudio realizado por Laan y cols. (2023), se evidencia la aplicación del método de Mann-Whitney U Test por parte de la autora para determinar la existencia de cambios significativos en el número de ataques de phishing. En particular, Laan clasifica los ataques de phishing en cinco categorías y, mediante el método mencionado, identifica cambios significativos en los cinco tipos de ataques de phishing durante el periodo de Covid-19 en comparación con el periodo pre-Covid-19.

Por lo que se puede concluir que 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 mali-

ciosos. 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.

3.3.2. Metodología

Descripción de los datos:

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 (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 los mismos.
- 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.

A continuación se presenta una muestra de las variables relevantes de la base de datos:

ID	name	start_date
3279.00	Unknown threat actor targeted Belgian pharmaceutical chain Goed on 18 March 2022	2022-03-18
3278.00	Unknown threat actors breached US lender Nations Direct Mortgage on 30 December 2023	2023-12-30
3277.00	Unknown hackers hijacked Instagram profile of Italian Prime Minister Giorgia Meloni	2023-03-17
3270.00	ShinyHunters obtained AT&T customer data in 2021 leaking over 70 million records on 17 March 2021	2021-01-01
3273.00	Unknown Threat Actor Hit Scottish NHS Dumfries & Galloway With Cyber Attack In March 2024	2022-03-01
3272.00	Unknown actors targeted Scranton School District in Pennsylvania with ransomware on 14 March 2020	2020-03-14

receiver_country	receiver_region	receiver_category	receiver_category_subcode
Belgium	EUROPE; EU(MS); NATO; WESTEU	Critical infrastructure	Health
United States	NATO; NORTHAM	Critical infrastructure	Finance
Italy	EUROPE; NATO; EU(MS)	State institutions / political system	Government / ministries
United States	NATO; NORTHAM	Critical infrastructure	Telecommunications
United Kingdom	EUROPE; NATO; NORTHEU	Critical infrastructure	Health
United States	NATO; NORTHAM	State institutions / political system; Education	Civil service / administration;

Se evidencia que los datos tienen el formato deseado, particularmente teniendo en cuenta el interés en el país y la región que recibe el ciberataque. La base de datos presenta pocos valores faltantes en estos dos aspectos, así como también en la columna de fecha. Considerando la escasez de valores faltantes y la completitud de los datos en relación con las variables de interés, y teniendo en cuenta la cantidad de datos, se determina que estos valores faltantes no representan un impedimento. Por lo tanto, se considera que esta base de datos es adecuada para el desarrollo de la investigación.

Con respecto al Análisis Exploratorio de datos, se tiene que: se valoran las siguientes tablas con estadísticos importantes sobre los ciberataques.

Cuadro 39: Medidas centralizadas y de dispersión

Medida	Valor
Total de Ciberataques	1300
Intensidad media	2.1
Nivel de intensidad máxima	5
Nivel de intensidad mínima	1.19
Grupos de ciberterroristas	525

Fuente: Elaboración propia con datos de Repositorio Europeo de Ciberincidentes.

Cuadro 40: Tabla de frecuencia según tipo de Ciberataque

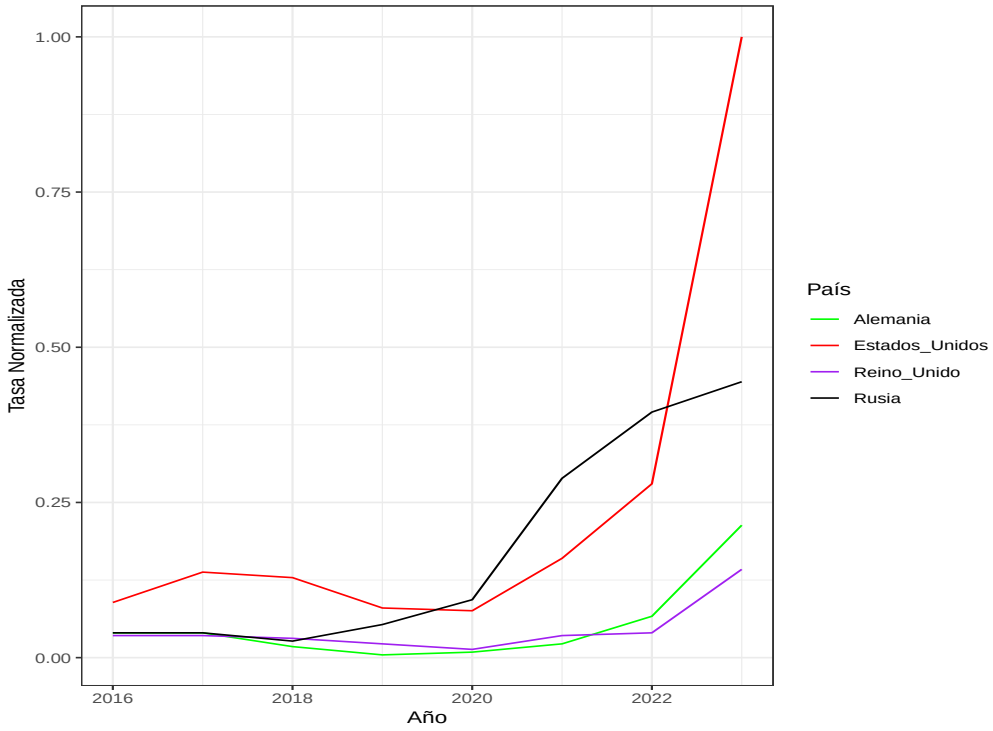
Tipo	Intensidad	Número de ciberataques
Ransomware	4	137
Secuestro con uso indebido	2.1	390
Robo de datos	2.20	303
Disrupción	2.96	270
Robo de datos y Voxing	2.73	90
Secuestro sin uso indebido	1.19	110

Fuente: Elaboración propia con datos de Repositorio Europeo de Ciberincidentes.

La base de datos refleja la cantidad de ciberataques de interés para el desarrollo de la investigación, correspondientes a los datos de 2016-2023 de Europa y Estados Unidos. Además, se incluye la intensidad media, mínima y máxima de los datos. Por otro lado, también se seccionan los datos por tipo de ciberataque para el periodo antes mencionado.

Por último, se presenta la siguiente gráfica, la cual ofrece una idea aproximada del comportamiento de los ciberataques en los países más afectados por este fenómeno.

Gráfico 11: Ciberataques normalizados por país



Fuente: Elaboración propia con datos del Repositorio Europeo de Ciberincidentes

En donde se esclarece como Estados Unidos es el país con mayor cantidad de ciberataques de forma general, seguido de Rusia.

Razón de ciberataques por usuario de internet(Anual y cuatrimestral)

En primera instancia, se notó que los datos proporcionados por la base de datos del repositorio europeo de ciberincidentes están en valores nominales. En otras palabras, para poder comparar estos datos, es necesario convertirlos a 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.

Cuadro 41: Número de usuarios de internet Europa (en millones) por año

Año	Usuarios de internet (millones)
2016	614.98
2017	659.63
2018	704.83
2019	727.56
2020	727.85
2021	743.60
2022	750.04
2023	726.02

Fuente: Elaboración propia con datos de Statista.

Cuadro 42: Número de usuarios de internet USA (en millones) por año

Año	Usuarios de internet (millones)
2016	282.10
2017	286.90
2018	286.90
2019	312.30
2020	288.10
2021	298.80
2022	307.20
2023	311.30

Fuente: Elaboración propia con datos de Statista

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.

Un ejemplo de este procedimiento se puede visualizar en:

Cuadro 43: Datos cuatrimestrales de ciberataques para USA (2016-2023)

Año	Cuatrimestre	Cantidad de ciberataques	Usuarios de internet	Razón de ciberataques por usuario
2016	1	15	94,033,333.33	1.59518E-07
2016	2	5	94,033,333.33	5.31726E-08
2016	3	5	94,033,333.33	5.31726E-08
2017	1	11	95,633,333.33	1.15023E-07
2017	2	7	95,633,333.33	7.31962E-08
2017	3	6	95,633,333.33	6.27396E-08
2018	1	12	95,633,333.33	1.25479E-07
2018	2	5	95,633,333.33	5.2283E-08
2018	3	10	95,633,333.33	1.04566E-07
2019	1	6	104,100,000	5.76369E-08
2019	2	8	104,100,000	7.68492E-08
2019	3	2	104,100,000	1.92123E-08
2020	1	9	96,033,333.33	9.37175E-08
2020	2	5	96,033,333.33	5.20653E-08
2020	3	1	96,033,333.33	1.04131E-08
2021	1	12	99,600,000	1.20482E-07
2021	2	9	99,600,000	9.03614E-08
2021	3	7	99,600,000	7.02811E-08
2022	1	10	102,400,000	9.76563E-08
2022	2	13	102,400,000	1.26953E-07
2022	3	23	102,400,000	2.24609E-07
2023	1	98	103,766,666.7	9.44427E-07
2023	2	53	103,766,666.7	5.10761E-07
2023	3	52	103,766,666.7	5.01124E-07

Fuente: Elaboración propia con datos de statista y repositorio europeo de ciberincidentes.

Prueba de Kolmogorov-Smirnov

Según DeGroot (2002), La prueba de Kolmogorov-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 de-

fine una 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 χ^2

Según Devore (2021), la prueba χ^2 se aplica para poder establecer una diferencia significativa entre la frecuencia esperada y las frecuencias observadas en categorías de tablas de contingencia. También, dado que la base de datos utilizada en la presente investigación involucra variables categóricas, la prueba χ^2 es una herramienta muy útil para conducir pruebas de hipótesis. (McHugh, 2013). De acuerdo a Zibran (2007), esta es una prueba estadística no paramétrica para determinar si dos o más clasificaciones de una muestra son o no independientes entre sí. La misma se realiza bajo las siguientes hipótesis (H.-Y. Kim, 2017) :

- H_0: Las variables son independientes (no hay asociación entre las variables)
- H_1: Las variables no son independientes (hay asociación entre las variables)

Se tienen las siguientes razones para escoger esta prueba estadística respecto de otras alternativas:

1. Es aplicable a variables categóricas, ya que todas las variables a estudiar de la base de datos son categóricas.
2. Es aplicable a tablas de contingencia de tamaños superiores al 2×2 y a muestras grandes. En el caso que se trata, la tabla de contingencia tiene tamaño 8×7,
3. Es sencilla de calcular e interpretar.

4. Con esta se puede obtener una medida de la magnitud de la asociación (de existir) entre dos variables categóricas. (McHugh, 2013).

Supuestos de la prueba χ^2

Naturalmente, es de suma importancia enunciar los supuestos de la prueba no paramétrica usada, principalmente para corroborar y sustentar una correcta aplicación de la misma en el contexto del presente trabajo. McHugh (2013) enumera seis supuestos de las pruebas χ^2, de los cuales los siguientes cuatro competen a la prueba de independencia:

1. Los datos en las celdas corresponden a frecuencias o conteos de casos, no a porcentajes u otro tipo de transformaciones de los datos.

2. Los niveles o categorías de las variables son mutuamente excluyentes, lo que se traduce en que una observación solo puede pertenecer a una celda. De esta manera, se evitan conteos dobles u otros inconvenientes o errores a nivel inferencial e interpretativo.

3. Hay solo dos variables involucradas, ambas categóricas y usualmente a nivel nominal, aunque también puede desarrollarse con categorías ordinales.

4. Al menos el 80 % de las celdas en la tabla de frecuencias esperadas son mayores o iguales a 5.

También, citando a Devore (2021) es relevante enunciar el teorema de Pearson y las condiciones de la bondad de ajuste de la prueba χ^2.

Teorema de Pearson χ^2

Cuando $H_0 : p_1 = p_{10}, ..., p_k = p_{k0}$ es verdadero, el estadístico

$$\chi^2 = \sum_{i=1}^{k} \frac{(N_i - np_{i0})^2}{np_{i0}}$$

tiene aproximadamente una distribución χ^2 con $k-1$ grados de libertad. Esta aproximación es determinada tal que $np_{i0} \geq 5$ para cada $i (i = 1, 2, ..., k)$

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.

Fundamentos de la Prueba de Signos

La prueba de signos se basa en la dirección de las diferencias entre pares de observaciones. Para cada par de observaciones (X_i, Y_i), se calcula la diferencia $D_i = X_i - Y_i$. Solo se consideran las diferencias distintas de cero:

1. **Signo positivo (+)**: Si $D_i > 0$
2. **Signo negativo (-)**: Si $D_i < 0$
3. **Se ignoran**: Si $D_i = 0$

La prueba de signos cuenta el número de signos positivos (S^+) y signos negativos (S^-).

Hipótesis

- **Hipótesis nula (H_0)**: Las medianas de las dos muestras emparejadas son iguales, es decir, no hay diferencia en la mediana de las dos muestras.
- **Hipótesis alternativa (H_1)**: Las medianas de las dos muestras emparejadas son diferentes.

Procedimiento de la Prueba

1. **Cálculo de las Diferencias**: Calcula las diferencias $D_i = X_i - Y_i$ para cada par de observaciones.
2. **Conteo de Signos**: Cuenta el número de diferencias positivas (S^+) y negativas (S^-).
3. **Cálculo de la Estadística de Prueba**: La estadística de prueba es el menor de los dos conteos ($S = \text{mín}(S^+, S^-)$).
4. **Distribución Binomial**: 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.
5. **Cálculo del Valor P**: Se calcula el valor p usando la distribución binomial:

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

Interpretación de Resultados

- **Valor p bajo** ($p < \alpha$): Rechazar la hipótesis nula. Hay evidencia significativa de que las medianas son diferentes.
- **Valor p alto** ($p \geq \alpha$): No se rechaza la hipótesis nula. No hay evidencia suficiente para decir que las medianas son diferentes.

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
- Al construir el estadístico se debe de aplicar una transformación usando las desviaciones absolutas respecto a la mediana, para la observación j-ésima del grupo i la transformación corresponde a:

$$Y_{ij} = \left| X_{ij} - \tilde{X}_i \right|$$

Supuestos de la prueba:

1. Las observaciones dentro de cada grupo son independientes:

$$\text{Cov}(X_{ij}, X_{ik}) = 0 \quad \text{para todo } j \neq k \text{ dentro de cada grupo } i$$

2. Las distribuciones de los grupos son similares en forma y ubicación, pero pueden diferir en su varianza:

$$X_{ij} \sim F(\mu_i, \sigma_i^2) \quad \text{donde } \epsilon_{ij} \sim F(0, \sigma_i^2)$$

donde:

- X_{ij} es la j-ésima observación en el grupo i
- μ_i es el parámetro de ubicación del grupo i
- σ_i^2 es la varianza del grupo i
- ϵ_{ij} son los errores aleatorios independientes con distribución F

Criterio de rechazo

Comparar el valor p obtenido con el nivel de significancia predefinido (generalmente $\alpha = 0{,}05$).

Valor $p < \alpha$: Rechazar la hipótesis nula, indicando una asociación significativa.

Valor $p \geq \alpha$: No rechazar la hipótesis nula, indicando falta de evidencia para una asociación significativa.

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 esta 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 las si-

guiente 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

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

Además, definimos:

p_{ij} como la probabilidad de que un individuo en la población pertenezca a la celda i, j con $i = 1, ..., R$; $j = 1, ...C$

p_{i+} como la probabilidad de que un individuo en la población se clasifique en la fila , calculado como

$$\mathbb{P}[\text{Individuo se clasifique en la fila} i] = \sum_{j=1}^{C} p_{ij}.$$

p_{+j} como la probabilidad de que un individuo en la población se clasifique en la columna j , calculado como

$$\mathbb{P}[\text{Individuo se clasifique en la columna} j] = \sum_{j=1}^{R} p_{ij}$$

Criterio de rechazo

Comparar el valor p obtenido con el nivel de significancia predefinido (generalmente $\alpha = 0{,}05$).

Valor $p < \alpha$: Rechazar la hipótesis nula, indicando una asociación significativa.

Valor $p \geq \alpha$: No rechazar la hipótesis nula, indicando falta de evidencia para una asociación significativa.

3.4. Resultados

En primera instancia, a partir de la ficha de resultados 25 y el gráfico número 12, se puede visualizar que la distribución de datos tanto para la región de Europa como para la de Estados Unidos no sigue

una distribución normal. Además, se observa que, en el caso de Estados Unidos, las diferencias entre las cantidades nominales 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 como el covid-19 genero una oleada de ciberataques en las regiones de Estados Unidos y Europa

Además de lo anterior, también se puede evidenciar en la ficha de resultados 26 y en los gráficos 13 y 14, que la gran 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 relativamente 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.

Por otra parte, los resultados obtenidos incluyen diversas herramientas estadísticas que solidifican la investigación. un ejemplo de esto se puede visualizar en 27 y en los gráficos 15 y 16, que son los resultados de la aplicación del cálculo de la razón de ciberataques anual y cuatrimestral para el periodo 2016-2023 en las regiones de interés, mediante la división entre la cantidad de ciberataques y la cantidad de usuarios de internet de forma anual y cuatrimestral. De esta manera, esta razón representa la proporción de ciberataques por persona que usa internet, permitiendo la comparación de los datos en diferentes momentos y proporcionando una visualización real del periodo con la mayor proporción de ciberataques.

Luego, mediante la formación de las poblaciones (Europa y Estados Unidos) y dos grupos de fechas (antes del COVID-19 (2017-2019) y durante y después del COVID-19 (2020-2023)), y al realizar la prueba Kolmogorov-Smirnov, se concluye lo evidenciado en 28 y 50, donde se refleja que existe 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.

Posteriormente, para implementar diferentes pruebas estadísticas, es importante determinar de manera confiable si las muestras anuales para ambas regiones presentan una distribución normal. Para

ello, se desarrolla una prueba de normalidad de los datos similar a la prueba Kolmogorov-Smirnov anterior, comparando cada muestra de datos del 2016-2023 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 calculo anterior en 51

Considerando lo anterior, lo siguiente es determinar la independencia de las variables. Dado que las muestras no son normales, se desarrolla una prueba no paramétrica para la independencia. Utilizando la misma segregación de poblaciones, se realiza la prueba Chi-cuadrado utilizando la razón de ciberataques por usuario de internet cuatrimestral. Se obtiene un valor de p mayor a 0.25 para ambas poblaciones52, por lo que no se puede rechazar la hipótesis nula de dependencia.De igual manera, Hawdon y cols. (2020), aplica la prueba chi-cuadrado para determinar la existencia de una relación, solo que en el caso de la autora trata de establecer relaciones entre los diferentes tipos de ciberataque.

Esto 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 paramétrica, que no depende fuertemente de la independencia. 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 (54). 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(54), 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 trato de determinar un cambios en las medias de 2 muestras, en este caso usando la prueba de signo se logro determinar la diferencia entre las medianas de las muestras.

Lo anterior permite concluir la existencia de una diferencia en la cantidad de ciberataques en las regiones de Europa y Estados Unidos durante el periodo del COVID-19 en comparación con el periodo pre-COVID-19. En particular, se puede concluir que existe un impacto en el número de ciberataques debido al COVID-19.

3.5. Parte de reflexión

3.5.1. Alcance investigación

A la luz del análisis realizado en las secciones anteriores, se determina el alcance real de la investigación. Por tanto, se actualiza la pregunta de investigación y los objetivos.

Pregunta de Investigación

¿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.

Objetivo General

Determinar 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.

Objetivos específicos

1. Examinar las características de los ciberataques más comunes y prominentes en Estados Unidos y Europa durante el periodo 2016-2023 mediante exploratorio de datos y revisión de literatura científica.

2. Hallar un modelo estadístico que facilite la visualización y el cálculo preciso del impacto del Covid-19 en el número de ciberataques en Europa y Estados Unidos, contrastando con el periodo prepandémico de 2016-2019.

3. Identificar las posibles implicaciones de la crisis sanitaria del Covid-19 en el número de ciberataques en Europa y Estados Unidos durante el periodo 2020-2023, en comparación con el periodo prepandemia, 2016-2019.

3.6. Actualización de UVE de Gowin

Conceptual

Conceptos:

- **Ciberataques:** acciones dirigidas contra sistemas de información para perjudicar.
- **Impacto:** efecto de una fuerza aplicada bruscamente.
- **Región:** porción de territorio determinada por características étnicas o circunstancias especiales.

Teorías y principios:

- Impacto del Covid-19 en el número de ciberataques.
- Modelos para determinar el impacto del Covid-19 en el número de ciberataques.

Objetivo general:

Determinar el impacto del Covid-19 en el número de ciberataques en Europa y Estados Unidos, en contraste con el periodo pre-pandemico 2016-2023.

Pregunta central:

¿Cuál es el impacto del Covid-19 en el número de ciberataques en Europa y Estados Unidos, en contraste con el periodo pre-pandemico 2016-2023?

Metodología

Metodología:

- **Razón o proporción:** Usado para determinar la razon o proporcion en la cantidad de ciberataque por usuario de internet. - **Prueba Kolmogorov-Smirnov:** Es una prueba de hipótesis que permite comparar las distribuciones de 2 muestras, así como para determinar independencia.
- **Prueba Chi-cuadrado:** Es una prueba de hipótesis no parametrica que permite determinar si existe relación entre las variables, usada comúnmente para determinar independencia.
-**Prueba Signos:** Es una prueba no parametrica de hipótesis que permite la comparación determinar si existe una diferencia entre las medianas de 2 muestras..

Objetivos específicos:

- Identificar las posibles implicaciones de la crisis sanitaria del Covid-19 en el número de ciberataques en Europa y Estados Unidos durante el periodo 2020-2023, en comparación con el periodo anterior a la pandemia, 2016-2019.
- Examinar las características de los ciberataques más comunes y prominentes en Estados Unidos y Europa durante el periodo 2016-2023 con el fin de identificar las posibles causas del aumento en su frecuencia debido al impacto del Covid-19.
- Hallar un modelo estadístico que facilite la visualización y el cálculo preciso del impacto del Covid-19 en el numero de ciberataques en Europa y Estados Unidos, contrastando con el periodo pre-pandémico de 2016-2019.

4. Bitácora 4

4.1. Parte de planificación

4.1.1. Fichas de resultados

Cuadro 44: Ficha de resultados 9

Encabezado	**Contenido**
Nombre de su hallazgo o resultado:	Prueba Exacta de Fisher para independencia
Resumen en una oración:	Se desarrolla prueba exacta de Fisher 53 y se determina que no hay evidencia significativa para rechazar la hipótesis nula y por tanto no se puede determinar que los datos son independientes
Principal característica:	Permite determinar mediante una tabla de contingencias 2x2 y con una pequeña cantidad de observaciones si los datos son independientes
Problemas o posibles desafíos:	Al necesitas una tabla 2x2 limita el uso de esta prueba y requiere ajustes los datos para poder aplicarla
Resumen en un párrafo:	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 exacta de Fisher en la razón de ciberataques por usuario de internet de forma cuatrimestral, se obtiene para ambas poblaciones un valor p muy alto y por tanto no se puede rechazar la hipótesis nula, lo cual lleva a que no se pueda concluir que los datos son independientes

Cuadro 45: Ficha de resultados 10

Encabezado	**Contenido**
Nombre de su hallazgo o resultado:	Prueba de Brown-Forsythe para comparar las varianzas entre 2 muestras
Resumen en una oración:	Se desarrolla la prueba de Brown-Forsythe 55 y se determina que para el nivel de significancia que se está considerando de 0.15, se rechaza la hipótesis nula de que las varianzas son iguales y por tanto hay una diferencia en la varianza de las muestras
Principal característica:	Prueba que no depende de normalidad y que es robusta con respecto a independencia
Problemas o posibles desafíos:	Considerar un nivel de significancia mayor a 0.05
Resumen en un párrafo:	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 Brown-Forsythe en la razón de ciberataques por usuario de internet cuatrimestral, se obtiene para la población de Europa un valor de p de 0.038. Considerando una significancia del 0.15, se rechaza la hipótesis nula de igualdad de varianzas y se determina la existencia de una diferencia con respecto a las varianzas. Ahora, con respecto a la población estadounidense, se obtiene un valor de p de 0.146. Por tanto, se puede rechazar la hipótesis nula y determinar que existe una diferencia con respecto a las varianzas.

Cuadro 46: Ficha de resultados 11

Encabezado	**Contenido**
Nombre de su hallazgo o resultado:	Existencia de un impacto en el número de ciberataques debido al Covid-19
Resumen en una oración:	Considerando las fichas de resultados referentes a la varianza y a la mediana, se logra determinar la existencia de un cambio en el número de ciberataques debido al Covid-19 con respecto a estos estadísticos para las regiones de Estados Unidos y Europa.
Principal característica:	Utilización de pruebas no paramétricas y robustas con respecto a la independencia
Problemas o posibles desafíos:	Se necesita considerar un nivel de significancia de 0.15 para poder rechazar las hipótesis nulas de ambas regiones para las pruebas de varianza y medianas
Resumen en un párrafo:	Desarrollando la prueba de signos y a prueba Brown-Forsythe se determina para cada región que tanto las medianas como las varianzas son diferentes para ambas muestras de datos 2016-2019 y 2020-2023, con lo cual se concluye que, al existir una diferencia significativa en ambos estadísticos, que el Covid-19 si genero un impacto con respecto al número de ciberataques tanto en Europa como también en Estados Unidos

Cuadro 47: Ficha de resultados 12

Encabezado	**Contenido**
Nombre de su hallazgo o resultado:	Imposibilidad de determinar independencia
Resumen en una oración:	Considerando la ficha de resultados Chi-cuadrado y a ficha de Prueba exacta de Fisher, se determina la imposibilidad de determinar independencia en las muestras
Principal característica:	Utilización de pruebas Chi-cuadrado y Prueba exacta de Fisher
Problemas o posibles desafíos:	Para pruebas posteriores, la dificultad de encontrar pruebas que no se vean afectadas en gran manera por la independencia y que sean robustas en este aspecto.
Resumen en un párrafo:	Desarrollando la prueba de Chi-cuadrado y la prueba exacta de Fisher, se determinó que para cada región las muestras de 2016-2019 y 2020-2023 de forma cuatrimestral, no pueden considerarse independientes. Esto se debe a que los valores de p resultantes son muy altos, esto considerando que el nivel d significancia elegido es de 0.15. Por lo tanto, no se logró establecer este aspecto.

4.2. Parte Escrita

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

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 Kolmogorov-Smirnov, Chi-cuadrado, 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 Kolmogorov-Smirnov busca identificar cambios en la distribución de los ataques cibernéticos, revelando patrones alterados bajo el impacto del COVID-19, así como determinar la normalidad de las muestras. La prueba de Chi-cuadrado se utiliza para determinar la asociación entre ciberataques en el periodo pandémico, contrastando con el periodo prepandémico. Además, para verificar esta asociación se utiliza la prueba exacta de Fisher. Además, la prueba de signos evalúa diferencias en las medianas de ataques antes y durante la pandemia, destacando un aumento significativo en la frecuencia y sofisticación 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. La investigación subraya la necesidad urgente de fortalecer las políticas de ciberseguridad para proteger infraestructuras críticas y datos sensibles en todos los ámbitos de la sociedad.

Palabras clave: Ciberataque, región, pandemia, impacto.

4.2.1. 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 etorno 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.

Identificar las posibles implicaciones de la crisis sanitaria del Covid-19 en el número de ciberataques en Europa y Estados Unidos durante el periodo 2020-2022, en comparación con el periodo prepandemia, 2016-2019. En primer lugar, 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.

Lallie y cols. (2021) también señalan que los ataques de phishing, ransomware y otras formas de cibercrimen se volvieron más comunes, con los atacantes adaptando sus estrategias para explotar el miedo y la incertidumbre generados por la pandemia. Las instituciones de salud y los organismos gubernamentales, en particular, fueron blanco de ataques debido a la urgencia y la sensibilidad de los datos manejados.

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.

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.

Laan y cols. (2023) ofrecen una perspectiva sobre cómo la frecuencia y el contenido de los ataques

de phishing se adaptaron durante la pandemia. Utilizando la teoría de actividades rutinarias y un modelo de elección racional del crimen, los autores argumentan que los cambios en las rutinas diarias y el aumento en la actividad en línea crearon oportunidades adicionales para los ataques de phishing. Laan y cols. (2023) muestra que los correos electrónicos de phishing se volvieron más personalizados y contextualmente relevantes, con temas relacionados con la Covid-19, como actualizaciones de salud, ayudas gubernamentales y teletrabajo. Estos correos electrónicos explotaron la necesidad urgente de información y la ansiedad generalizada, aumentando la probabilidad de éxito de los ataques.

Además, el estudio realizado por Hawdon y cols. (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 otro lado, en el estudio realizado por Laan y cols. (2023), se evidencia la aplicación del método de Mann-Whitney U Test para determinar la existencia de cambios significativos en el número de ataques de phishing. En particular, Laan clasifica los ataques de phishing en cinco categorías y, mediante el método mencionado, identifica cambios significativos en los cinco tipos de ataques de phishing durante el periodo de Covid-19 en comparación con el periodo pre-Covid-19.

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.

- Prueba Kolmogorov-Smirnov

- Prueba χ^2
- Prueba exacta de Fisher
- Prueba de signos de Wilcoxon.
- Prueba Brown-Forsythe

4.2.2. 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 (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.

A continuación se presenta una muestra de la base de datos:

ID	name	start_date
3279.00	Unknown threat actor targeted Belgian pharmaceutical chain Goed on 18 March 2022	2022-03-18
3278.00	Unknown threat actors breached US lender Nations Direct Mortgage on 30 December 2023	2023-12-30
3277.00	Unknown hackers hijacked Instagram profile of Italian Prime Minister Giorgia Meloni	2023-03-17
3270.00	ShinyHunters obtained AT&T customer data in 2021 leaking over 70 million records on 17 March 2021	2021-01-01
3273.00	Unknown Threat Actor Hit Scottish NHS Dumfries & Galloway With Cyber Attack In March 2024	2022-03-01
3272.00	Unknown actors targeted Scranton School District in Pennsylvania with ransomware on 14 March 2020	2020-03-14

inclusion_criteria	incident_type	receiver_name
Attack on critical infrastructure target(s)	Disruption; Hijacking with Misuse	Goed
Attack on critical infrastructure target(s)	Data theft; Hijacking with Misuse	Nations Direct Mortgage
Attack on (inter alia) political target(s), not politicized	Hijacking with Misuse	Giorgia Meloni
Attack on critical infrastructure target(s)	Data theft & Doxing; Hijacking with Misuse	AT&T
Attack on critical infrastructure target(s)	Data theft; Hijacking with Misuse	NHS Dumfries and Galloway
Attack on (inter alia) political target(s), not politicized	Disruption; Hijacking with Misuse; Ransomware	Scranton School District

receiver_country	receiver_region	receiver_category	receiver_category_subcode
Belgium	EUROPE; EU(MS); NATO; WESTEU	Critical infrastructure	Health
United States	NATO; NORTHAM	Critical infrastructure	Finance
Italy	EUROPE; NATO; EU(MS)	State institutions / political system	Government / ministries
United States	NATO; NORTHAM	Critical infrastructure	Telecommunications
United Kingdom	EUROPE; NATO; NORTHEU	Critical infrastructure	Health
United States	NATO; NORTHAM	State institutions / political system; Education	Civil service / administration;

initiator_country	attributing_country	unweighted_cyber_intensity	target_multiplier	weighted_cyber_intensity
Not available	Not available	0.00	Moderate - high political importance	3.00
Not available	Not available	3.00	Moderate - high political importance	3.00
Not available	Not available	2.00	Moderate - high political importance	2.00
Not available	Not available	3.00	Moderate - high political importance	3.00
Not available	Not available	3.00	Moderate - high political importance	3.00
Not available	Not available	4.00	Moderate - high political importance	4.00

Fuente: Elaboración propia con datos de Repositorio Europeo del Ciber incidentes

Pruebas estadística

Razón de ciberataques por usuario de internet(Anual y cuatrimestral)

En primera instancia, se notó que los datos proporcionados por la base de datos del repositorio europeo de ciberincidentes están en valores nominales. En otras palabras, para poder comparar estos datos, es necesario convertirlos a 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.

Cuadro 48: Número de usuarios de internet Europa (en millones) por año

Año	Usuarios de internet (millones)
2016	614.98
2017	659.63
2018	704.83
2019	727.56
2020	727.85
2021	743.60
2022	750.04
2023	726.02

Fuente: Elaboración propia con datos de Statista.

Cuadro 49: Número de usuarios de internet USA (en millones) por año

Año	Usuarios de internet (millones)
2016	282.10
2017	286.90
2018	286.90
2019	312.30
2020	288.10
2021	298.80
2022	307.20
2023	311.30

Fuente: Elaboración propia con datos de Statista

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.

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

Prueba χ^2

Según Devore (2021), la prueba χ^2 se aplica para poder establecer una diferencia significativa entre la frecuencia esperada y las frecuencias observadas en categorías de tablas de contingencia. También, dado que la base de datos utilizada en la presente investigación involucra variables categóricas, la prue-

ba χ^2 es una herramienta muy útil para conducir pruebas de hipótesis. (McHugh, 2013). De acuerdo con Zibran (2007), esta es una prueba estadística no paramétrica para determinar si dos o más clasificaciones de una muestra son o no independientes entre sí. La misma se realiza bajo las siguientes hipótesis (H.-Y. Kim, 2017) :

- H_0: Las variables son independientes (no hay asociación entre las variables)
- H_1: Las variables no son independientes (hay asociación entre las variables)

Se tienen las siguientes razones para escoger esta prueba estadística respecto de otras alternativas:

1. Es aplicable a variables categóricas, ya que todas las variables a estudiar de la base de datos son categóricas.
2. Es aplicable a tablas de contingencia de tamaños superiores al 2×2 y a muestras grandes. En el caso que se trata, la tabla de contingencia tiene tamaño 8×7,
3. Es sencilla de calcular e interpretar.
4. Con esta se puede obtener una medida de la magnitud de la asociación (de existir) entre dos variables categóricas. (McHugh, 2013).

Supuestos de la prueba χ^2

Naturalmente, es de suma importancia enunciar los supuestos de la prueba no paramétrica usada, principalmente para corroborar y sustentar una correcta aplicación de esta en el contexto del presente trabajo. McHugh (2013) enumera seis supuestos de las pruebas χ^2, de los cuales los siguientes cuatro competen a la prueba de independencia:

1. Los datos en las celdas corresponden a frecuencias o conteos de casos, no a porcentajes u otro tipo de transformaciones de los datos.
2. Los niveles o categorías de las variables son mutuamente excluyentes, lo que se traduce en que una observación solo puede pertenecer a una celda. De esta manera, se evitan conteos dobles u otros inconvenientes o errores a nivel inferencial e interpretativo.
3. Hay solo dos variables involucradas, ambas categóricas y usualmente a nivel nominal, aunque también puede desarrollarse con categorías ordinales.
4. Al menos el 80 % de las celdas en la tabla de frecuencias esperadas son mayores o iguales a 5.

También, citando a Devore (2021) es relevante enunciar el teorema de Pearson y las condiciones de la bondad de ajuste de la prueba χ^2.

Teorema de Pearson χ^2

Cuando $H_0 : p_1 = p_{10}, ..., p_k = p_{k0}$ es verdadero, el estadístico

$$\chi^2 = \sum_{i=1}^{k} \frac{(N_i - np_{i0})^2}{np_{i0}}$$

tiene aproximadamente una distribución χ^2 con $k-1$ grados de libertad. Esta aproximación es determinada tal que $np_{i0} \geq 5$ para cada $i (i = 1, 2, ..., k)$

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.

Fundamentos de la Prueba de Signos

La prueba de signos se basa en la dirección de las diferencias entre pares de observaciones. Para cada par de observaciones (X_i, Y_i), se calcula la diferencia $D_i = X_i - Y_i$. Solo se consideran las diferencias distintas de cero:

1. **Signo positivo (+)**: Si $D_i > 0$
2. **Signo negativo (-)**: Si $D_i < 0$
3. **Se ignoran**: Si $D_i = 0$

La prueba de signos cuenta el número de signos positivos (S^+) y signos negativos (S^-).

Hipótesis

- **Hipótesis nula (H_0)**: Las medianas de las dos muestras emparejadas son iguales, es decir, no hay diferencia en la mediana de las dos muestras.
- **Hipótesis alternativa (H_1)**: Las medianas de las dos muestras emparejadas son diferentes.

Procedimiento de la Prueba

1. **Cálculo de las Diferencias**: Calcula las diferencias $D_i = X_i - Y_i$ para cada par de observaciones.

2. **Conteo de Signos**: Cuenta el número de diferencias positivas (S^+) y negativas (S^-).

3. **Cálculo de la Estadística de Prueba**: La estadística de prueba es el menor de los dos conteos ($S = \text{mín}(S^+, S^-)$).

4. **Distribución Binomial**: 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.

5. **Cálculo del Valor P**: 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

- Al construir el estadístico se debe de aplicar una transformación usando las desviaciones absolutas respecto a la mediana, para la observación j-ésima del grupo i la transformación corresponde a:

$$Y_{ij} = \left| X_{ij} - \tilde{X}_i \right|$$

Supuestos de la prueba:

1. Las observaciones dentro de cada grupo son independientes:

$$\text{Cov}(X_{ij}, X_{ik}) = 0 \quad \text{para todo } j \neq k \text{ dentro de cada grupo } i$$

2. Las distribuciones de los grupos son similares en forma y ubicación, pero pueden diferir en su varianza:

$$X_{ij} \sim F(\mu_i, \sigma_i^2) \quad \text{donde } \epsilon_{ij} \sim F(0, \sigma_i^2)$$

donde:

- X_{ij} es la j-ésima observación en el grupo i
- μ_i es el parámetro de ubicación del grupo i
- σ_i^2 es la varianza del grupo i
- ϵ_{ij} son los errores aleatorios independientes con distribución F

Criterio de rechazo

Comparar el valor p obtenido con el nivel de significancia predefinido (generalmente $\alpha = 0{,}05$).
Valor $p < \alpha$: Rechazar la hipótesis nula, indicando una asociación significativa.
Valor $p \geq \alpha$: No rechazar la hipótesis nula, indicando falta de evidencia para una asociación significativa.

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

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

Además, definimos:

p_{ij} como la probabilidad de que un individuo en la población pertenezca a la celda i, j con $i = 1, ..., R$; $j = 1, ...C$

p_{i+} como la probabilidad de que un individuo en la población se clasifique en la fila , calculado como

$$\mathbb{P}[\text{Individuo se clasifique en la fila} i] = \sum_{j=1}^{C} p_{ij}.$$

p_{+j} como la probabilidad de que un individuo en la población se clasifique en la columna j , calculado como

$$\mathbb{P}[\text{Individuo se clasifique en la columna} j] = \sum_{j=1}^{R} p_{ij}$$

Criterio de rechazo

Comparar el valor p obtenido con el nivel de significancia predefinido (generalmente $\alpha = 0{,}05$).

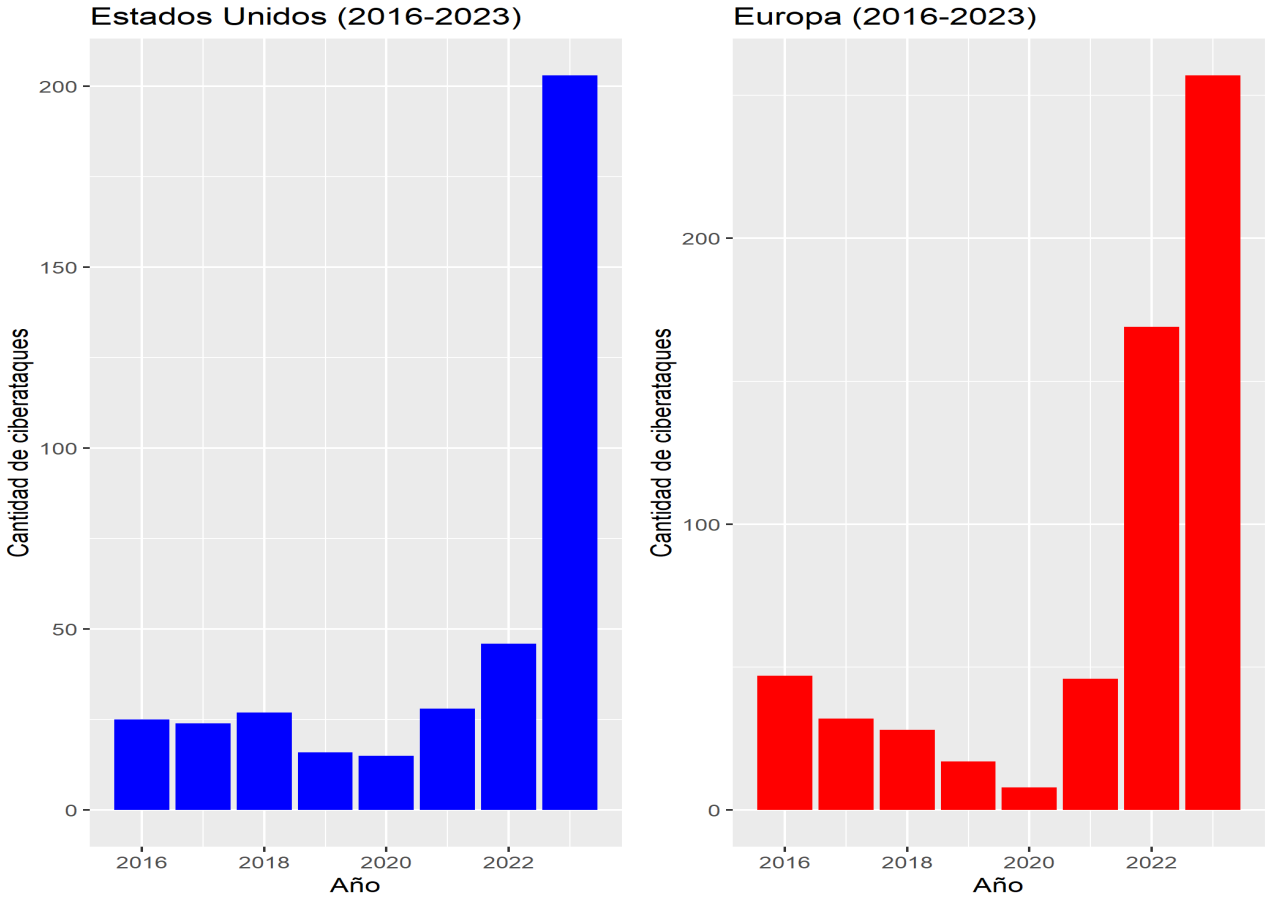
Valor $p < \alpha$: Rechazar la hipótesis nula, indicando una asociación significativa.

Valor $p \geq \alpha$: No rechazar la hipótesis nula, indicando falta de evidencia para una asociación significativa.

4.2.3. Resultados

En primera instancia, a partir de la ficha de resultados 25 y el gráfico número 12, se puede visualizar que la distribución de datos tanto para la región de Europa como para la de Estados Unidos no sigue una distribución normal. Además, se observa que, en el caso de Estados Unidos, las diferencias entre las cantidades nominales 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.

Gráfico 12: Histogramas de Ciberataques por Año

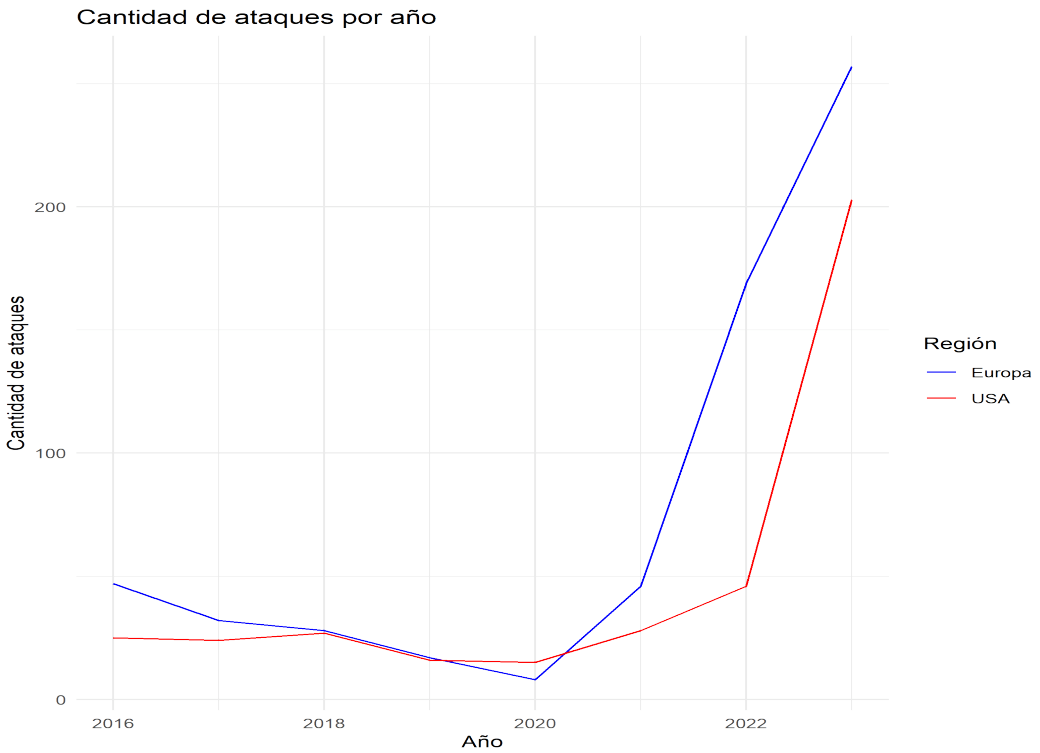


Fuente: Elaboración propia con datos del Repositorio Europeo de Ciberincidentes

Además de lo anterior, también se puede evidenciar en la ficha de resultados 26 y en los gráficos 13 y 14, que la gran cantidad de ciberataques presentes en Estados Unidos es notable. A pesar de que

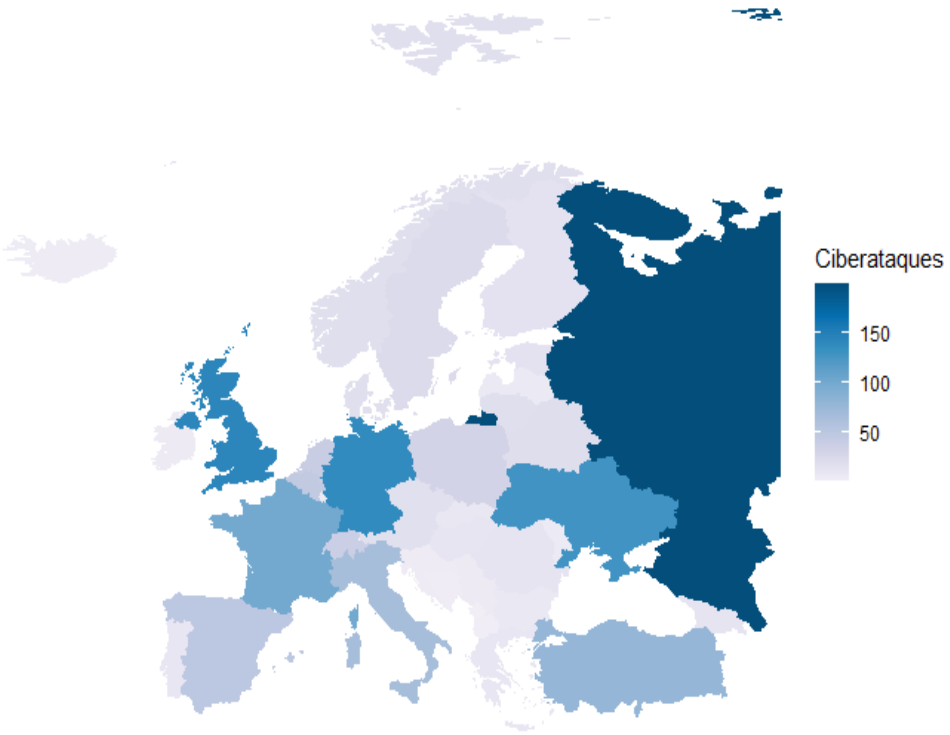
solo se trata de un país, el número de ciberataques presenta una diferencia relativamente pequeña en comparación con toda la región de Europa, lo que refleja claramente la magnitud de los casos en Estados Unidos en comparación con otras regiones. Además, la cantidad de ciberataques en Estados Unidos 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.

Gráfico 13: Gráfico de lineas de Ciberataques por Año



Fuente: Elaboración propia con datos del Repositorio Europeo de Ciberincidentes

Gráfico 14: Países de Europa por cantidad de ciberataques recibidos



Fuente: Elaboración propia con datos del Repositorio Europeo de Ciberincidentes

Por otra parte, los resultados obtenidos incluyen diversas herramientas estadísticas que solidifican la investigación. Un ejemplo de esto se puede visualizar en 27 y en los gráficos 15 y 16, que son los resultados de la aplicación del cálculo de la razón de ciberataques anual y cuatrimestral para el periodo 2016-2023 en las regiones de interés, mediante la división entre la cantidad de ciberataques y la cantidad de usuarios de internet de forma anual y cuatrimestral. De esta manera, esta razón representa la proporción de ciberataques por persona que usa internet, permitiendo la comparación de los datos en diferentes momentos y proporcionando una visualización real del periodo con la mayor proporción de ciberataques..

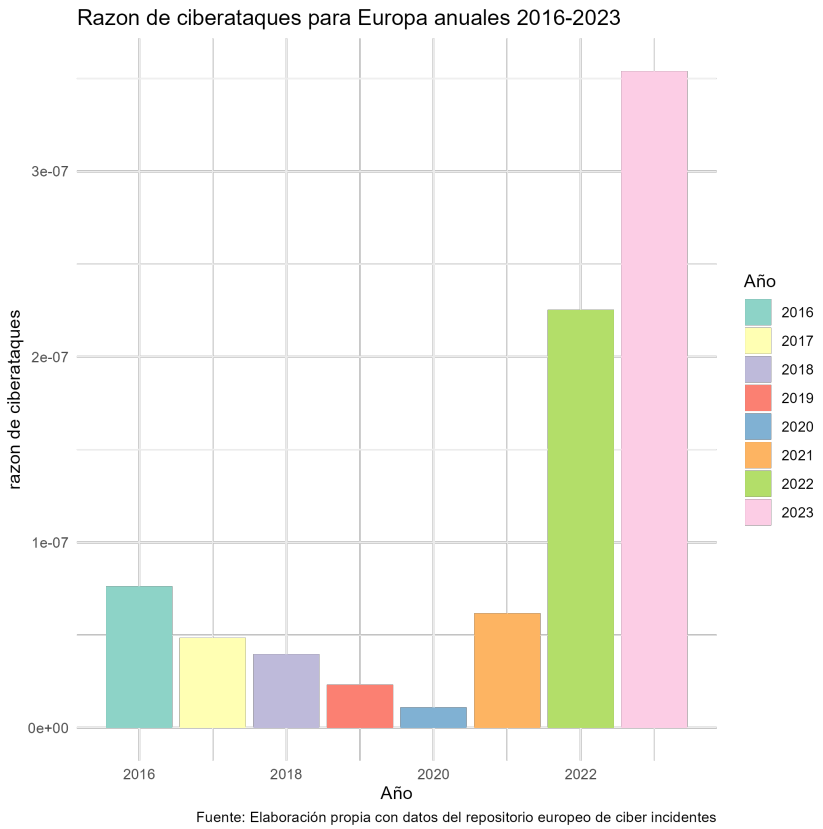


Gráfico 15

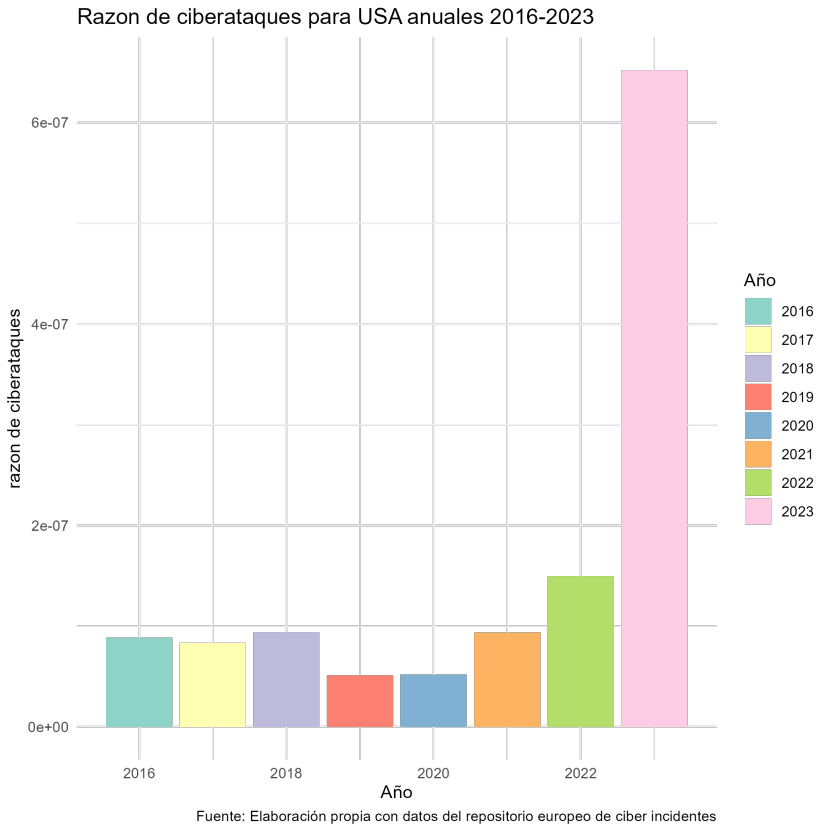


Gráfico 16

Luego, mediante la formación de las poblaciones (Europa y Estados Unidos) y dos grupos de fechas (antes del COVID-19 (2017-2019) y durante y después del COVID-19 (2020-2023)), y al realizar la prueba Kolmogorov-Smirnov, se concluye lo evidenciado en 28 y 50, donde se refleja que existe 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 50: 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 anuales para ambas regiones presentan una distribución normal. Para ello, se desarrolla una prueba de normalidad de los datos similar a la prueba Kolmogorov-Smirnov anterior, comparando cada muestra de datos del 2016-2023 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 51

Cuadro 51: 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, lo siguiente es determinar la independencia de las variables. Dado que las muestras no son normales, se desarrolla una prueba no paramétrica para la independencia. Utilizando la misma segregación de poblaciones, se realiza la prueba Chi-cuadrado utilizando la razón de ciberataques por usuarios de internet cuatrimestral. Se obtiene un valor de p mayor a 0.25 para ambas poblaciones52, por lo que no se puede rechazar la hipótesis nula de dependencia. Sumado a lo anterior

y debido a la cantidad de observaciones presente en la base de datos, también se decide desarrollar una prueba exacta de Fiches para independencia, esto ya que, por la cantidad de observaciones, la prueba chi-cuadrado podría generar algún error, con lo aplicando la prueba exacta de Fisher 53, se obtienen valores p muy altos y, por tanto, se concluye, al igual que con la prueba chi-cuadrado, no se puede rechazar la hipótesis nula y, por tanto, no se determina que las muestras son independientes. De igual manera, Hawdon y cols. (2020), aplica la prueba chi-cuadrado para determinar la existencia de una relación, solo que en el caso de la autora trata de establecer relaciones entre los diferentes tipos de ciberataque.

Cuadro 52: Resultados de la prueba Chi-cuadrado para Europa y USA cuatrimestral para 2 muestras (2016-2019) y (2020-2023)

País	**Estadístico X-squared**	**Grados de libertad (df)**	**Valor p**
Europa	96	88	0.2625
USA	120	110	0.2421

Fuente: Elaboración propia con datos del repositorio europeo de Ciber incidentes.

Cuadro 53: Resultados de la Prueba exacta de Fisher en los periodos 2017-2019 y 2020-2023 para USA y Europa

Resultados de la prueba exacta de Fisher cuatrimestral para Europa y USA para 2 muestras (2016-2019) y (2020-2023)		
Periodo 2016-2019	Periodo 2020-2023	p-value
Europa	Europa	1
USA	USA	1

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

Esto 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 (54). 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(54), 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 54: 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 55, 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 55: 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.

4.2.4. 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, Chi-cuadrado y la prueba exacta de Fisher. Los resultados iniciales indicaron que las muestras no siguen una distribución normal, no son independientes y, lo más importante, 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.

5. 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.

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.

Devore, C., Berk. (2021). *Modern mathematical statistics with application*. Springer. Descargado de https://link.springer.com/book/10.1007/978-1-4614-0391-3

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.

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

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 cri-

me. *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.

6. 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`