



1859

UNL

Universidad
Nacional
de Loja

Universidad Nacional de Loja

Facultad de la Educación, El Arte y la Comunicación

Carrera de Pedagogía de las Ciencias Experimentales

**Uso de datos abiertos para la enseñanza de la Estadística
Descriptiva en el nivel secundario con Python**

Trabajo de Integración Curricular,
previo a la obtención del título de
Licenciado en Pedagogía de las
Matemáticas y la Física.

AUTOR:

Jean Carlo Prado Vélez

DIRECTOR:

Fabricio Vladimir Vines Vines

**Loja – Ecuador
2024**



unl

Universidad
Nacional
de Loja

Certificación

Sistema de Información Académico
Administrativo y Financiero - SIAAF

CERTIFICADO DE CULMINACIÓN Y APROBACIÓN DEL TRABAJO DE INTEGRACIÓN CURRICULAR

Yo, **VINCES VINCES FABRICIO VLADIMIR**, director del Trabajo de Integración Curricular denominado **Uso de datos abiertos para la enseñanza de la Estadística Descriptiva en el nivel secundario con Python**, perteneciente al estudiante **JEAN CARLO PRADO VELEZ**, con cédula de identidad N° **1105484404**.

Certifico:

Que luego de haber dirigido el **Trabajo de Integración Curricular**, habiendo realizado una revisión exhaustiva para prevenir y eliminar cualquier forma de plagio, garantizando la debida honestidad académica, se encuentra concluido, aprobado y está en condiciones para ser presentado ante las instancias correspondientes.

Es lo que puedo certificar en honor a la verdad, a fin de que, de así considerarlo pertinente, el/la señor/a docente de la asignatura de **Integración Curricular**, proceda al registro del mismo en el Sistema de Gestión Académico como parte de los requisitos de acreditación de la Unidad de Integración Curricular del mencionado estudiante.

Loja, 31 de Julio de 2024



FABRICIO VLADIMIR
VINCES VINCES

F)

DIRECTOR DE TRABAJO DE INTEGRACIÓN
CURRICULAR



Certificado TIC/TT.: UNL-2024-001546

1/1

Educamos para **Transformar**

Autoría

Yo, **Jean Carlo Prado Vélez**, declaro ser autor del presente Trabajo de Integración Curricular y eximo expresamente a la Universidad Nacional de Loja y a sus representantes jurídicos de posibles reclamos o acciones legales, por el contenido del mismo. Adicionalmente acepto y autorizo a la Universidad Nacional de Loja, la publicación de mi Trabajo de Integración Curricular en el Repositorio Digital Institucional - Biblioteca Virtual.

Firma:

A handwritten signature in blue ink that reads "Jean Prado". The signature is written in a cursive style with a large initial 'J' and 'P'.

Cédula de identidad: 1105484404

Fecha: 21 de octubre de 2024

Correo electrónico: jean.prado@unl.edu.ec

Teléfono: +593 967825265

Carta de autorización por parte del autor para consulta, reproducción parcial o total y/o publicación electrónica del texto completo, del Trabajo de Integración Curricular.

Yo, **Jean Carlo Prado Vélez**, declaro ser autor del Trabajo de Integración Curricular, denominado: **Uso de datos abiertos para la enseñanza de la Estadística Descriptiva en el nivel secundario con Python**, como requisito para optar el título de **Licenciado en Pedagogía de las Matemáticas y la Física**, autorizo al Sistema Bibliotecario de la Universidad Nacional de Loja, para que, con fines académicos, muestre la producción intelectual de la Universidad, a través de la visibilidad de su contenido de la siguiente manera en el Repositorio Institucional.

Los usuarios pueden consultar el contenido de este trabajo en el Repositorio Institucional, en las redes de información del país y del exterior con las cuales tenga convenio la Universidad.

La Universidad Nacional de Loja, no se responsabiliza por el plagio o copia del Trabajo de Integración Curricular que realice un tercero.

Para constancia de esta autorización, en la ciudad de Loja, a los veintiún días del mes de octubre de dos mil veinticuatro, firma el autor.



Firma: _____

Autor: Jean Carlo Prado Vélez

Cédula: 1105484404

Dirección: Loja, Ciudadela Alegría

Correo electrónico: jean.prado@unl.edu.ec

Teléfono: +593 967825265

DATOS COMPLEMENTARIOS:

Director del Trabajo de Integración Curricular: Lic. Fabricio Vladimir Vines Vines Mg.
Sc

Dedicatoria

Quiero dedicar este trabajo a mis Padres y Abuelas, quienes con sus consejos me han estado motivándome cada día a seguir adelante en cada paso que doy. A mis hermanos que han contribuido de manera significativa en mi crecimiento personal y académico. Y a todas las personas que me apoyaron constantemente y me brindaron sus consejos para hacer de mí una mejor persona, y a mis amigos, en especial a mis compañeros: Rosaura Jiménez, Valeria Cabrera, Anthony Becerra y Alex Matailo que me han apoyado durante todo el proceso académico, por su paciencia, confianza, palabras de ánimo y estímulo y amistad incondicional.

Jean Carlo Prado Vélez

Agradecimiento

Agradezco a mi Madre y Abuela por su apoyo incondicional, por jamás descuidarse y despreocuparse de mí, ni un solo segundo, cada paso en este trabajo ha sido impulsado por su confianza en mí, y a mis queridos amigos que fueron como mi segunda familia, gracias por estar siempre ahí compartiendo este viaje inolvidable.

De igual forma, expreso mi más sincero agradecimiento a mi director de tesis, Fabricio Vladimir Vincés, por su valiosa orientación y apoyo constante a lo largo de este trabajo de investigación los aportes de sus conocimientos, experiencias y sugerencias permitieron el avance del presente trabajo. Al rector de la Unidad Educativa “TCRN. Lauro Guerrero”, Juan Luna Rengel. Mg.Sc., por brindarme su valiosa colaboración para la ejecución de la presente investigación.

Jean Carlo Prado Vélez

Índice de contenidos

| | |
|--|-------------|
| Portada | i |
| Autoría | iii |
| Dedicatoria | v |
| Agradecimiento | vi |
| Índice de contenidos | vii |
| Índice de tablas: | viii |
| Índice de figuras: | viii |
| Índice de anexos: | viii |
| 1. Título | 1 |
| 2. Resumen | 2 |
| Abstract | 3 |
| 3. Introducción | 4 |
| 4. Marco Teórico | 6 |
| La Estadística en la sociedad..... | 6 |
| Estadística..... | 6 |
| <i>¿Qué es la estadística?</i> | 6 |
| <i>Estadística Descriptiva</i> | 7 |
| <i>Estadística Inferencial</i> | 8 |
| La enseñanza de la Estadística Descriptiva en la educación | 9 |
| <i>Paradigmas de la enseñanza de la estadística descriptiva</i> | 10 |
| <i>Dificultades en la enseñanza de la Estadística Descriptiva</i> | 11 |
| Estrategias para la enseñanza de la Estadística Descriptiva | 14 |
| <i>La tecnología y el aprendizaje de la Estadística</i> | 15 |
| <i>Beneficios de la tecnología en la enseñanza de la Estadística</i> | 16 |
| <i>Softwares en la enseñanza de la Estadística</i> | 17 |
| <i>Uso de datos abiertos</i> | 22 |
| <i>Datos abiertos Ecuador</i> | 23 |
| Implementación de datos abiertos en la enseñanza de la Estadística con Python..... | 24 |
| Estadística en el sistema educativo ecuatoriano..... | 28 |
| 5. Metodología | 30 |
| 6. Resultados | 33 |
| 7. Discusión | 39 |
| 8. Conclusiones | 41 |

| | |
|---------------------------------|-----------|
| 9. Recomendaciones | 42 |
| 10. Bibliografía | 43 |
| 11. Anexos | 52 |

Índice de tablas:

| | |
|--|-----------|
| Tabla 1. Softwares estadísticos más conocidos..... | 18 |
| Tabla 2. Ecuaciones de búsqueda utilizadas | 31 |
| Tabla 3. Procesos de enseñanza de la Estadística Descriptiva en nivel secundario utilizando datos abiertos con Python..... | 33 |

Índice de figuras:

| | |
|--|-----------|
| Figura 1 ¿Qué gráfico elegir?..... | 8 |
| Figura 2 Proceso de análisis de datos estadísticos. | 9 |
| Figura 3. Proceso de enseñanza de la Estadística Descriptiva con datos abiertos y Python 36 | |
| Figura 4 Preguntas enfocadas al objetivo específico 2 | 37 |

Índice de anexos:

| | |
|---|------------|
| Anexo 1. Propuesta Didáctica..... | 52 |
| Anexo 2. Oficio de Apertura de la Institución y certificado de la ejecución de experiencia áulica. 104 | |
| Anexo 3. Bitácora de búsqueda..... | 106 |
| Anexo 4. Certificado de traducción del resumen..... | 109 |

1. Título

Uso de datos abiertos para la enseñanza de la Estadística Descriptiva en el nivel
secundario con Python

2. Resumen

En el siglo XXI, la Estadística es importante para la toma de decisiones informadas y su enseñanza en la educación es esencial, especialmente cuando se emplean herramientas digitales como Python para manejar datos reales. Por ello, esta investigación busca analizar el proceso de enseñanza de la Estadística Descriptiva en nivel secundario utilizando datos abiertos gestionados con Python, para esto se realizó un proceso riguroso de revisión documental sistemática, en artículos, y capítulos de libros. La búsqueda de información se efectuó en bases de datos científicas y motores de búsqueda, las técnicas para selección y organización se realizaron mediante bitácoras de búsqueda, fichas bibliográficas y de contenido. Con base en esa teoría se elaboró 6 talleres abordados como clases en una institución Fiscomisional de la ciudad de Loja, donde se abordó temáticas de Estadística Descriptiva haciendo uso de datos abiertos y Python, siguiendo una secuencia de Inicio (3 talleres), Desarrollo (2 talleres) y Cierre (1 taller), finalmente se aplicó un instrumento para medir el nivel de satisfacción de los estudiantes. Como resultados se obtuvo un proceso de enseñanza adecuado para esta disciplina y un grado de satisfacción alto. Gracias a esto se concluyó que los estudiantes evidencian quedar satisfechos, lo que implica que el uso de la secuencia de las 3 fases y la inclusión de una herramienta digital genera cierta predisposición en el alumno para aprender Estadística Descriptiva.

Palabras clave: *Herramientas digitales, Software estadístico, Estadística aplicada*

Abstract

In the 21st century, Statistics is important for informed decision making and its teaching in education is essential, especially when digital tools such as Python are used to handle real data. Therefore, this research seeks to analyze the teaching process of Descriptive Statistics at the secondary level using open data managed with Python, for this a rigorous process of systematic documentary review was carried out, in articles, and book chapters. The search for information was carried out in scientific databases and search engines, the techniques for selection and organization were carried out by means of search logs, bibliographic and content files. Based on this theory, 6 workshops were developed as classes in a Fiscomisional institution in the city of Loja, where Descriptive Statistics topics were addressed using open data and Python, following a sequence of beginning (3 workshops), Development (2 workshops) and Closing (1 workshop), finally an instrument was applied to measure the level of satisfaction of the students. The results showed an adequate teaching process for this discipline and a high degree of satisfaction. Thanks to this, it was concluded that the students were satisfied, which implies that the use of the sequence of the 3 phases and the inclusion in the student to learn Descriptive Statistics

Keywords: Digital tools, Statistical software, Applied statistics

3. Introducción

La enseñanza de la Estadística constituye hoy en día una herramienta esencial que permite el procesamiento de datos, aporta al desarrollo del razonamiento estadístico y otorga a los estudiantes herramientas e ideas para que puedan responder inteligentemente a la amplia información del mundo actual. Además, es de gran importancia, ya que permite adaptar a las personas a la realidad social, permitiendo comprender el mundo en el que viven, es decir, analizar los datos y la información que nos topamos diariamente, como la lectura de la información nutricional, los censos, entender resultados electorales, entre otros miles de ejemplos cotidianos que hace fundamental el conocimiento de la estadística.

La Estadística Descriptiva es una de las ramas que presenta más dificultades, ya sea por errores al momento de realizar cálculos para obtener resultados estadísticos o por la poca comprensión de los conceptos, los datos no logran organizarlos de manera correcta y por ello no pueden realizar la interpretación de los mismos, algunos estudiantes ignoran el estudio de algunos temas puesto que consideran que la información adquirida no sea de gran utilidad (Cox et al., 2022 y Arce et al., 2019). Sin embargo, con el desarrollo de las nuevas tecnologías de la información y su incorporación en la educación, ha permitido que la enseñanza de la estadística se lleve a cabo de manera adecuada, puesto que el docente ahora cuenta con herramientas tecnológicas más actualizadas, como los softwares y paquetes estadísticos.

Con lo antes mencionado, se estableció la categoría conceptual: Uso de datos abiertos para la enseñanza de la Estadística Descriptiva en el nivel secundario con Python, y se planteó como problema de investigación: ¿Cómo mejora el proceso de enseñanza de la Estadística Descriptiva en el nivel secundario utilizando datos abiertos gestionados con Python?, con el siguiente objetivo general: Analizar el proceso de enseñanza de la Estadística Descriptiva en nivel secundario utilizando datos abiertos gestionados con Python.

Para dar respuesta a esta pregunta de investigación se plantearon los siguientes objetivos específicos: indagar los diferentes procesos de enseñanza de la Estadística Descriptiva en nivel secundario utilizando datos abiertos gestionados con Python; identificar el grado de satisfacción de los estudiantes del nivel secundaria al estudiar Estadística Descriptiva utilizando datos abiertos y Python, y diseñar una propuesta didáctica que describa el proceso de enseñanza de la Estadística Descriptiva en nivel secundario utilizando datos abiertos tratados con Python.

Cabe indicar que el estudio de este tema es de gran relevancia ya que mediante ella se puede analizar e interpretar procedimientos, datos y gráficos estadísticos que son importantes

para realizar una gran variedad de investigaciones. La enseñanza de los contenidos de esta disciplina debe estar rodeada de diversos recursos tanto didácticos como tecnológicos que faciliten el proceso de aprendizaje de la misma.

Esta investigación sigue los lineamientos y estructura preestablecida en el Reglamento de Régimen Académico de la Universidad Nacional de Loja, en la que se incluyen los elementos siguientes: título, Uso de datos abiertos para la enseñanza de la Estadística Descriptiva en el nivel secundario con Python; resumen, se detalla las ideas más esenciales del trabajo además de las palabras claves; introducción, muestra de forma clara el contenido del informe; marco teórico, revisión bibliográfica y documental sistemática que permitió sustentar científicamente el trabajo; metodología, se detalla el proceso llevado a cabo en el uso de métodos, técnicas y herramientas; resultados, descripción de datos cualitativos y cuantitativos que responde a los objetivos de la investigación; conclusiones, información relevante del trabajo; recomendaciones, se derivan de las conclusiones; bibliografía, bajo el estilo de normas APA 7; y, anexos, se detalla información complementaria al marco teórico, tal como la propuesta didáctica y la bitácora de búsqueda.

4. Marco Teórico

La Estadística en la sociedad

Durante el diario vivir el individuo se enfrenta a grandes cantidades de datos e información recibida a través de diferentes medios de comunicación, ante esto, es de vital importancia poseer un pensamiento crítico capaz de poder realizar interpretaciones y análisis para tomar decisiones, así como diferenciar información relevante de aquella que no ha sido informada de manera correcta. Según lo expuesto Begg (1997, como se citó en Batanero, 2000) señala que “la estadística es un buen vehículo para alcanzar las capacidades de comunicación, tratamiento de la información, resolución de problemas, uso de ordenadores y trabajo cooperativo y en grupo, a las que se da gran importancia en los nuevos currículos” (p. 2). Además, Ruiz (2015) menciona que, en la actualidad, debido al desarrollo de la sociedad de la información, el auge de la tecnología, y a la necesidad de poder interpretar la extensa información disponible, ha provocado que la Estadística cobre mayor importancia.

La Estadística ha estado presente en la vida cotidiana desde siempre y durante el transcurso del tiempo ha demostrado tener una aplicabilidad positiva llegando a mejorar la calidad de vida de la sociedad, por esta razón ha adquirido un papel importante en la sociedad pues gracias a esta disciplina se logran interpretar fenómenos a partir del procesamiento de datos. Por lo que Ponteville (2014) menciona que en el desarrollo de la sociedad actual la estadística cumple un papel primordial, puesto que brinda herramientas que dan la posibilidad de describir situaciones de incertidumbre en análisis científicos, sociales económicos actuales. A su vez aborda aspectos tales como el análisis de la variabilidad, las relaciones entre variables, el diseño de estudios, experimentos y la mejora de predicciones.

La estadística es aplicable en diversos campos, sirve para la toma de mejores decisiones a partir de la comprensión de las fuentes de variación y de la detección de patrones y relaciones en los datos obtenidos, en la resolución de problemas en sectores empresariales y políticos, en el desarrollo científico e incluso en la vida diaria. Por estos motivos es considerado un pilar fundamental para interpretar y analizar la creciente cantidad de datos que nos rodea, dando fundamento a las decisiones, basadas en evidencia.

Estadística

¿Qué es la estadística?

La Estadística comúnmente suele ser considerada como una relación de datos numéricos presentada de forma ordenada y sistemática. Sin embargo, podemos decir que la Estadística es aquella ciencia que estudia el cómo se debe utilizar la información ya sea para

ayudar en la toma de decisiones o para explicar condiciones regulares o irregulares de algún fenómeno o estudio aplicado. Según Ross (2007) “La Estadística es el arte de aprender a partir de los datos. Está relacionada con la recopilación de datos, su descripción subsiguiente y su análisis, lo que nos lleva a extraer conclusiones” (p.3).

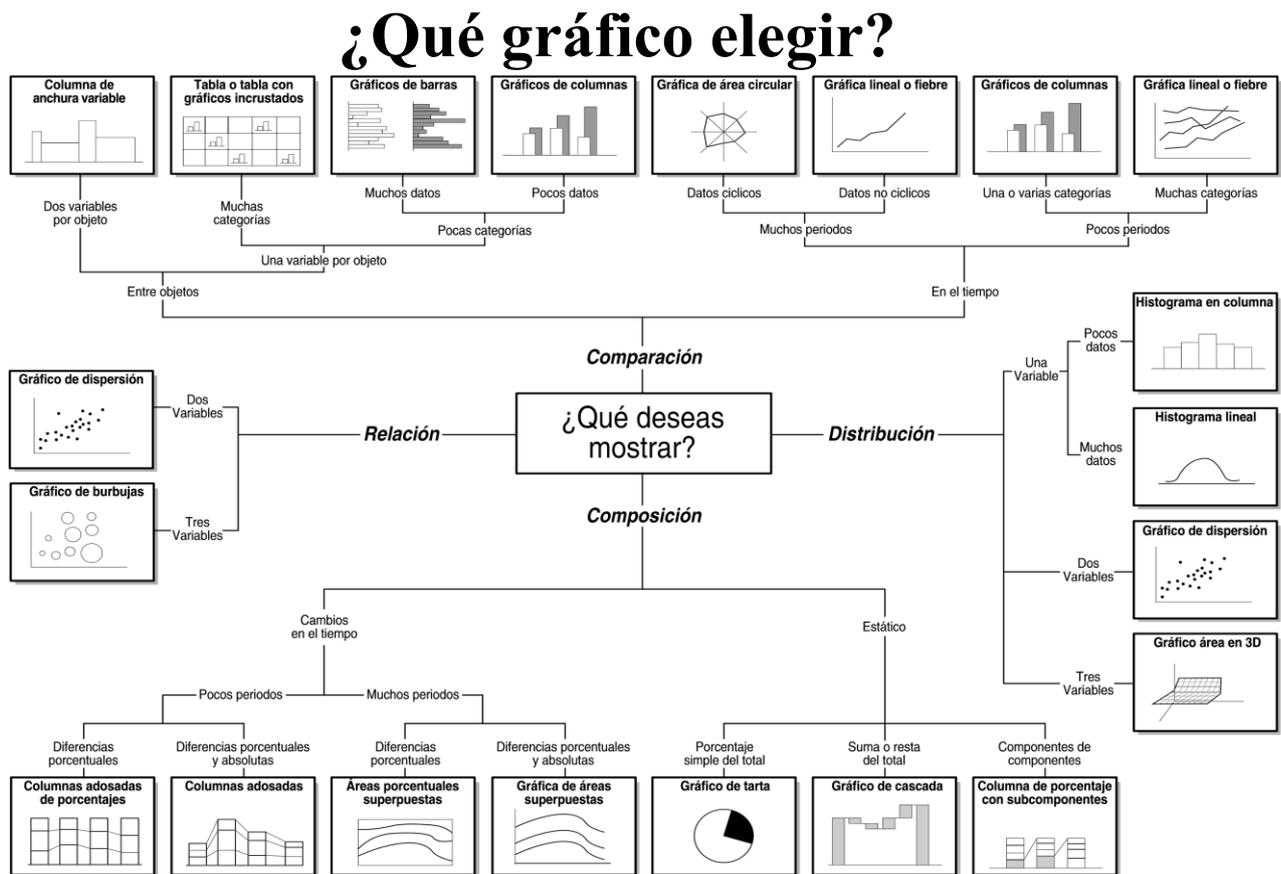
La Estadística como disciplina proporciona una metodología basada en la Matemática, con el fin de obtener, recopilar, procesar y presentar datos procedentes de algún estudio de interés para obtener conclusiones. Esta disciplina de gran importancia se divide en varias ramas las cuales desempeñan un papel crucial en el análisis e interpretación de los datos en diversos campos, cada una de estas ramas tiene su propio conjunto de métodos y técnicas eficaces para resolver problemas estadísticos. La estadística se divide en dos ramas principales: Estadística Descriptiva y Estadística Inferencial. A continuación, procederemos a describir cada una de estas ramas.

Estadística Descriptiva.

Esta rama se ocupa principalmente del resumen y descripción de los datos, aquí los métodos estadísticos se encargan principalmente de la recopilación, organización, presentación, análisis e interpretación de estos, Barreto (2012) define a la Estadística descriptiva como los métodos que implican la recopilación, caracterización y presentación de un conjunto de datos con el propósito de describir sus características. En relación con este tema, Mesa y Caicedo (2020) indican que su finalidad es “describir las características principales de una muestra, lo cual se puede realizar mediante cuadros, gráficos o índices” (p. 12).

Asimismo, Rendón et al. (2016) hacen referencia a que la Estadística Descriptiva formula sugerencias sobre como resumir información en cuadros y tablas, además de ello busca integrar y dar coherencia a los resultados de un fenómeno o trabajo de manera ordenada, sencilla y clara con el fin de que puedan ser interpretadas tanto por otros investigadores como por lectores. Es por esto importante, que la representación de los datos debe ser adecuada y eligiendo la gráfica correcta, para ello se toma en cuenta la Figura 1, de Abala (2009) y el gráfico original se lo puede encontrar, vectorizado, en [7. Charts - Extreme Presentation.](#)

Figura 1 ¿Qué gráfico elegir?



Nota. La figura muestra una serie de técnicas visuales, las cuales están clasificadas dependiendo de la información que se requiera presentar, ya sea cualitativa o cuantitativa.

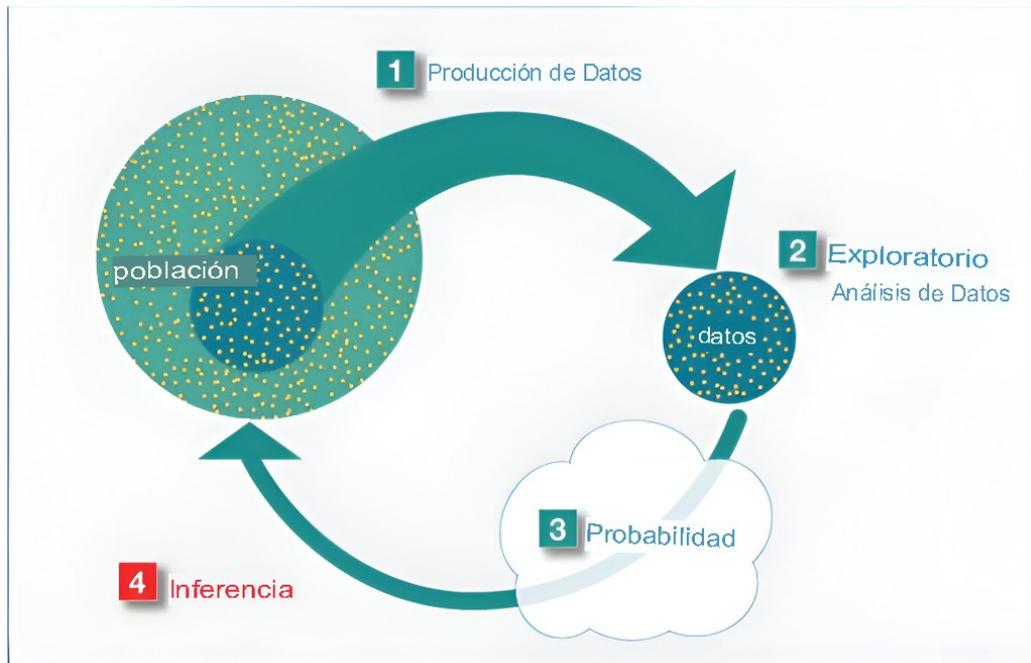
Estadística Inferencial.

Esta rama de la estadística permite hacer un proceso inductivo para inferir sobre una medida estadística con base en observaciones de una muestra seleccionada en el estudio. Según Mesa y Caicedo (2020) “La Estadística Inferencial se apoya en el Cálculo de Probabilidades y usa los resultados de la Estadística Descriptiva con el fin de generalizar y aplicar los conceptos a la población” (p. 12). Es decir, que se ocupa de brindar los métodos necesarios para emitir conclusiones validas sobre las características de la población mediante juicios probabilísticos, por tanto, la teoría de la probabilidad, así como todo lo relativo a variables aleatorias discretas y continuas, son las bases para llevar a cabo la estadística inferencial.

Las diversas aplicaciones de la Estadística han hecho que se convierta en un método efectivo para detallar los resultados de investigaciones políticas, económicas, sociales, entre otros. Además, sirve como herramienta para el análisis de los datos y a partir de ello tomar decisiones (Figura 2), por ello, además de tenerla como ciencia se hizo importante y necesario considerarla como materia de enseñanza en la sociedad.

Figura 2

Proceso de análisis de datos estadísticos.



Nota. La figura muestra el proceso para llevar a cabo de manera ordenada y sistemática el análisis de los datos estadísticos, que va desde su producción hasta la inferencia que se hace para sacar conclusiones basadas en los datos. Adaptado de LibreTexts (2024).

La enseñanza de la Estadística Descriptiva en la educación

La enseñanza de la Estadística constituye hoy en día una herramienta esencial que permite el procesamiento de datos, incluye actividades con diversos fines como lo económico, social, político entre otras que son áreas muy destacables dentro del mundo moderno, puesto que permite la inclusión en la toma de decisiones, así como la proyección de elementos de juicio, evaluaciones y las correspondientes investigaciones para solucionar todo tipo de problemas.

La enseñanza de esta disciplina en los últimos años ha ido tomando cada vez más importancia en la educación, puesto que presenta una mayor integración en los diversos niveles académicos ya que no solo aporta al desarrollo del razonamiento estadístico, sino que otorga a los estudiantes herramientas, ideas y disposiciones para que puedan responder inteligentemente a la amplia información del mundo actual (Estrella, 2017). De modo similar, Wild et al. (2018) menciona que la educación Estadística en el mundo actual está orientada precisamente a equipar a los estudiantes de marcos conceptuales, promoviendo distintas formas de pensar,

razonar estadísticamente y al desarrollo de habilidades prácticas de esta disciplina con el fin de que puedan navegar con éxito por las turbulentas aguas de la sociedad actual.

La enseñanza de la estadística se considera una nueva disciplina en constante consolidación y expansión, dado que está logrando avances considerables, incluso mayores que otras ramas de la matemática, tanto en amplitud como en profundidad (Batanero, 2018). Es por ello que, Salinas y Mayén (2016) afirman que aprender estadística se ha vuelto algo indispensable para el ciudadano, puesto que le permite analizar, interpretar y tomar decisiones basadas en la información.

La aplicación de la Estadística es muy importante en diversas áreas como salud, educación, economía, finanzas, ingeniería, política, otros, debido, según Chaves (2016) a su potencial para la recolección y análisis de información que se genera dentro de las mismas con el fin de argumentar valoraciones e investigaciones que a futuro van a generar nuevos conocimientos. Así mismo, Cuétara et al. (2016) menciona que la enseñanza de la Estadística en el mundo actual se lleva a cabo en un entorno en el que se destaca las ideas de nuevos enfoques que logren romper las cadenas de diferentes concepciones, estrategias y por el reconocimiento de los límites que presentan estudiantes y docentes antes, durante y después de finalizar sus estudios.

Por otro lado, Chaves (2012) menciona que la mayoría de propuestas didácticas de enseñanza de la Estadística Descriptiva, están inmersas en un modelo constructivista, el cual busca generar un aprendizaje significativo, es decir, que su objetivo es que los estudiantes sean en gran medida los protagonistas al momento de adquirir los conocimientos estadísticos mediante la participación activa, exploración y reflexión.

Estos autores consideran que esta disciplina requiere un tipo diferente de pensamiento con respecto a las matemáticas ya que consideran que los datos no son simplemente números, sino que son números que están inmersos dentro de un contexto, tal contexto da significado a los datos por ello es importante tenerlo en cuenta al momento de llevar a cabo la enseñanza de la Estadística.

Paradigmas de la enseñanza de la estadística descriptiva

En los últimos años han destacado varios paradigmas, los cuales han servido como sustento teórico y guía para la enseñanza de esta disciplina. Históricamente, tales paradigmas son considerados producto de posiciones platónicas en las que atribuyen a la mente humana un papel muy importante al momento de llevar a cabo cualquier actividad en la vida cotidiana. Dentro de este pensamiento, el conocimiento llegó a ser visto como “algo” que se descubre y solo puede demostrarse si se almacena en la mente misma.

Batanero (2001) hace referencia a que el paradigma constructivista puede ser potencialmente útil al momento de llevar a cabo la enseñanza de la Estadística, puesto que considera que el conocimiento es desarrollado de manera activa por el individuo y no recibido desde el entorno. Además, menciona que el sujeto está en constante adaptación al mundo que le rodea, y al momento de que una nueva idea se presenta sobre otras ya existentes se presenta un "conflicto cognitivo" el cual se soluciona a través de un proceso de "equilibración", es decir, que la posibilidad de aprender depende del conocimiento adquirido y del desarrollo intelectual del alumnado.

Por otro lado, Cuevas e Ibáñez (2008) consideran que el paradigma cognitivo dentro de la estadística hace alusión a que el estudiante haga algo con la información que le es brindada por el docente, para ello sugiere manipular los datos con el fin de estar en condiciones adecuadas para llevar a cabo la construcción de su propio conocimiento. Es decir que, en este paradigma el aprendizaje implica una asimilación en donde el estudiante no solo se va a limitar a la adquisición de los conocimientos, sino que, con la ayuda del docente, va a construir usando sus conocimientos previos.

Para que la Estadística tenga valor y sea relevante para los estudiantes, Calderwood (2002) considera al paradigma pragmático importante dentro de esta disciplina puesto que, permite a los alumnos ver cómo se aplica lo que es enseñado, es decir, este paradigma permite que los estudiantes valoren lo que se les enseña. Sin embargo, para lograr esto es necesario que tanto docentes como estudiantes se hagan preguntas tales como: ¿Es esto algo que los investigadores y quienes toman decisiones realmente utilizan?, ¿Es esto algo que los estudiantes harán en su vida educativa, personal o profesional?, ya que si lo que están aprendiendo no se aplica en situaciones de la vida diaria, entonces no será de interés o valor para ellos.

Dificultades en la enseñanza de la Estadística Descriptiva.

En la actualidad, en la mayoría de las áreas de estudio se enseña al menos un curso de Estadística. Sin embargo, sigue siendo común que los estudiantes presenten dificultades para comprender lo que es enseñado en esta disciplina. Dichas dificultades pueden depender de diversos factores entre ellos: el área en que se estudia, la metodología utilizada por los docentes al momento de impartir sus clases o incluso el grado de preparación que los estudiantes poseen.

Según Cox et al. (2022) menciona que la Estadística Descriptiva es una de las ramas que más presenta dificultades, ya sea por errores al momento de realizar cálculos para obtener resultados estadísticos o por la poca comprensión de los conceptos, un ejemplo de ello es que algunos estudiantes no logran distinguir la media de la mediana, también a la hora de presentar

datos, estos no logran organizarlos de manera correcta y por ello se presenta otra dificultad que es la interpretación de los mismos. Agregado a lo anterior, en estudios recientes como es el caso de Arce et al. (2019), menciona que algunos estudiantes ignoran el estudio de algunos temas como lo es las medidas de dispersión puesto que consideran que la información adquirida no sea de gran utilidad además de confundir conceptos como varianza con variación.

En este mismo sentido, Ramón y Vilchez (2020), mencionan que la mayoría de los docentes suelen utilizar algunos métodos y estrategias para llevar a cabo la enseñanza de conceptos y procesos estadísticos, pero de acorde a los conocimientos que ellos poseen o tomando en cuenta una bibliografía predeterminada con ejemplos descontextualizados. En muchos casos abordan una enseñanza repetitiva, siguiendo los mismos patrones en los que fueron formados, ignorando las necesidades de los estudiantes e incluso no poseen conocimientos necesarios para hacer uso de la tecnología como medio y recurso para el procesamiento de la información. En la misma línea Alfaro et al. (2015) hace referencia a que al momento de llevar a cabo la enseñanza de la Estadística han llegado a observar que los docentes muchas de las veces enseñan con base a reproducción de determinados algoritmos, lo que puede llegar a provocar en los estudiantes mecanización de los procesos de un estudio estadístico lo que conlleva a dejar de lado el análisis e interpretación de la información.

Los docentes no enseñan la estadística de una manera significativa sino que es llevada a cabo de manera mecánica y memorístico es decir, que debido a esto no logran desarrollar ningún tipo de pensamiento, más bien lo que obtienen como resultado son estudiantes incapaces de razonar, analizar, no son críticos, no pueden trabajar en contextos, no toman en cuenta los intereses y necesidades de los estudiantes, no hay una mejora en sus estrategias didácticas por lo tanto sus enseñanzas carecen de motivación y atención (Moreno y Aguerre, 2015).

Otras dificultades que están presentes en la enseñanza de Estadística Descriptiva según León (2020) es el enfoque calculista que de ninguna manera permite captar la real esencia de la Estadística en su verdadera dimensión como la ciencia de los datos y la incertidumbre por el contrario, dicho enfoque hace que los estudiantes no le encuentren sentido a esta disciplina. Así mismo, Guajardo (2016) menciona que es muy común encontrar estudiantes que desde un principio demuestran actitudes negativas dirigidas hacia la Estadística, lo que conlleva a derivar en un impedimento tanto en la enseñanza como en el aprendizaje.

Por otro lado, Zúñiga y Díaz (2017) indican que el docente no hace uso de situaciones problema dentro del aula clases, solo se basa en la parte teórica y no la aplican, esto provoca en los estudiantes la dificultad de no poder aplicar lo aprendido en situaciones del mundo real.

En los tiempos actuales la enseñanza de la estadística requiere ser realizada en espacios de aprendizaje activo en donde se involucren a los estudiantes en problemas o situaciones reales, estudios estadísticos ya que esto permitirá una mejor comprensión de los contenidos y una mejora en cómo interpretar la información.

Igualmente, para Espinoza y Sánchez (2014) otra dificultad en la enseñanza y aprendizaje de los contenidos de Estadística Descriptiva, es que los estudiantes no tienen nociones básicas tales como; comprender lo que es un dato y cómo se recopilan, el tipo de variables, medidas de tendencia central, medidas de dispersión, y gráficos estadísticos, los cuales son necesarios puesto que son conocimientos bases que les sirven para que logren llevar lo aprendido a lo práctico, ya que de nada les serviría enseñar más contenidos si no logran comprender lo básico de esta disciplina.

El hecho de que la estadística se incluyan de una forma oficial en el currículo no significa que necesariamente se enseñen. Algunos estudios como el de Azcárate (2006), menciona que muchos profesores no se sienten cómodos con estas materias, es por ello que la dejan como último tema y cuando es posible la omiten. Del mismo modo, Ruiz (2015) hace referencia a que es muy frecuente que los docentes con tal de cumplir con el currículo llegan a dar las clases de forma superficial encajadas con cálculos extensos mediante un sin número de fórmulas, dejando de lado el razonamiento y la relación que tiene esta disciplina con otras que puedan ser de gran relevancia ya que gracias a estas se puede dar mayor importancia al aprendizaje con el fin de evitar desfases en conocimientos que perjudiquen su continuación en etapas posteriores.

Para Azcárate (2006) uno de los factores que más ha marcado el déficit de conocimientos estadísticos en los estudiantes es la actitud que tiene el docente hacia la enseñanza puesto que algunos de ellos no poseen con la formación adecuada sobre la materia en sí y tampoco han sido capacitados en ello. Es por estas razones que muchas de las veces a esta disciplina no la toman con tanta relevancia, incluso al igual que las matemáticas Ballesteros (2013) menciona que los estudiantes la perciben como un conocimiento complicado, aburrido, amenazador en lugar de verla como una oportunidad ya que es un aprendizaje instrumental.

Estrategias para la enseñanza de la Estadística Descriptiva.

Cabe indicar que tanto la enseñanza como el aprendizaje de la Estadística Descriptiva pueden resultar más efectivos cuando se aplican diversas estrategias que involucran a los estudiantes de manera activa y fomenten la comprensión de conceptos estadísticos, es por esta razón que los docentes han buscado adaptar diferentes estrategias con el fin de despertar el interés en los estudiantes hacia la Estadística, algunas de las estrategias que se utilizan es la contextualización de problemas en donde se da una relación de los conceptos estadísticos con situaciones de la vida cotidiana, logrando así que los estudiantes vean a esta disciplina más relevante y significativa. Es más, Wistuba (2014) destaca que los modelos de enseñanza para la estadística se utilizan con fines educativos con el propósito de acercar a los estudiantes al conocimiento estadístico de una forma mucho más práctica y llamativa, con el fin de garantizar la replicación de lo aprendido.

Así mismo, Gordon (1995) afirma que “la importancia de que la enseñanza requiere ser construida sobre la base de experiencias personales de quien est[á] aprendiendo, tratar de conectar conceptos abstractos con experiencias personales, analogías, símiles y metáforas, pueden ser herramientas instruccionales muy útiles” (p. 200). Bajo esta misma óptica, Santoyo (2022) plantea la necesidad de que la enseñanza de la estadística involucre al estudiante en la resolución de problemas reales, a partir de la ejecución de proyectos que incluyan problemas estadísticos, así como a través del análisis de datos a partir de su contexto. Es aquí donde entra una nueva estrategia de innovación en el aprendizaje pues se trata de otro tipo de enseñanza que se ha estado promoviendo y es aquel basado en proyectos, el cual está dando mejores resultados en los cursos de estadística (Behar et al, 2013). En esta enseñanza son los estudiantes son quienes realizan sus propios proyectos, teniendo una vinculación más con la realidad y comprendiendo los conocimientos de una manera más práctica y eficiente.

Según Rosabal et al. (2021) las estrategias didácticas deben ofrecer una amplia variedad de actividades de aprendizaje que proporcione una diversidad en las formas en la que los alumnos aprenden, cumpliendo así con los objetivos y contenidos planteados. Estos autores consideran que la motivación es una estrategia pedagógica considerada de gran impacto en la enseñanza de la estadística, puesto que la definen como un sentimiento que crea necesidades, que dirige y sostiene el comportamiento de las personas, que determina el fin o la continuación de una actividad.

Además de ello, Alfaro et al. (2015) consideran a las TIC (Tecnologías de la Información y la Comunicación) una estrategia didáctica en la enseñanza y aprendizaje de la Estadística Descriptiva, ya que estas son uno de los principales medios para poder acceder a

cualquier tipo de información, por lo tanto, es considerado un importante recurso, en el cual los estudiantes aprenden de forma autónoma, de tal manera que facilitan la realización de cálculos, gráficos y análisis a través del uso de software, simuladores, hojas de cálculo, laboratorios en línea, entre otros.

La enseñanza de la estadística se ha reflejado un conjunto de cambios dados por la aparición de nuevas herramientas, estas han logrado facilitar el cálculo de operaciones en donde los profesores han incorporado al aula de clases estas innovaciones tecnológicas. Dicho de otro modo, la estadística va más allá de resolver procedimientos matemáticos y repetirlos, actualmente se trata de reflexionar, interpretar resultados operacionales mediante los nuevos medios tecnológicos, suministrando así una nueva estrategia didáctica para enseñar y aprender estadística. La tecnología puede apoyar tanto a los docentes en la enseñanza de la estadística como a los estudiantes para que logren reforzar y alcanzar mejores resultados de aprendizaje.

La tecnología y el aprendizaje de la Estadística.

De acuerdo con Cardona (2023), la necesidad de vincular las TIC en los procesos de enseñanza y aprendizaje, es responsabilidad que recae en el quehacer educativo de los docentes. Las nuevas apariciones de tecnologías, metodologías y herramientas, deben ser consideradas como un impulso para lograr capacitar, actualizar e innovar a los educadores en estas áreas del conocimiento. En la misma línea, los autores Behar y Ojeada (2016) mencionan que la tecnología debe ser considerada como un valioso instrumento de apoyo, que con el pasar del tiempo tendrá un mejor uso por parte del profesor, puesto que, puede utilizarla y ver donde es potencialmente más efectiva.

Así mismo, Quevedo et al. (2015) mencionan que las tecnologías ofrecen varios recursos para la construcción del conocimiento, sin embargo, el introducir estas herramientas requiere de cierto tiempo para poder llevar a cabo una planeación consciente y estructurada que asegure un material adecuado para el usuario. Las intervenciones y los cambios en el sistema educativo están estrechamente ligados a múltiples pruebas de innovaciones que se realizan en contextos controlados para verificar el grado de logro de los objetivos planteados. Es por ello que se consideran a las TIC dentro de la enseñanza y el aprendizaje de la Estadística fundamentales, por lo que deben ser utilizadas de manera correcta con el fin de desarrollar un aprendizaje significativo, en donde los estudiantes puedan ver a la Estadística como una asignatura emocionante e interesante de aprender y no como aquella a la que simplemente deben aprobar por aprobar.

Según Belfiori (2014) las TIC son aquellos recursos tecnológicos informáticos y telecomunicaciones que están dirigidos a mejorar los procesos de información y comunicación, estos al ser llevados a cabo dentro de la enseñanza han contribuido a mejorar los procesos de creación de contenidos multimedia, nuevos escenarios de apertura e inclusive entornos colaborativos. Este mismo autor menciona que existen una amplia gama de recursos TIC disponibles para llevar a cabo la enseñanza de la Estadística, desde los específicos tales como SimStat, BioStat, StadisS (herramientas de análisis estadístico de datos) y otros, hasta los más comunes en los ordenadores, celulares y tablets que poseen los alumnos tales como Excel y Geogebra. Al introducir la tecnología en las planificaciones para la enseñanza de la estadística no es con la intención de solo resolver ejercicios mecánicos, sino también para mejorar la adquisición de los conocimientos mediante distintas formas de obtener información, para luego poder compartir los resultados de las actividades que se han llevado a cabo.

De la misma manera, Rodríguez et al. (2016) mencionan que las TIC nos ofrecen una variedad de recursos de apoyo para los procesos de enseñanza como lo son: material didáctico, softwares interactivos, entornos virtuales, internet, blogs, wikis, webquest, foros, chat, mensajerías, videos conferencias y otros canales de comunicación y manejo de información. Tales recursos logran facilitar el desarrollo de la creatividad, permiten la innovación y creación de entornos de trabajo colaborativo y además de ello promover un aprendizaje significativo, activo y flexible para el estudiante.

El hecho de utilizar recursos tecnológicos no quiere decir que el papel de los estudiantes va a pasar a un segundo plano, sino que va a sufrir una modificación, es decir, va a pasar de realizar diversos cálculos a tomar decisiones y hacer las respectivas interpretaciones con los resultados obtenidos. Es por ello que Zuluaga y Villa (2018) mencionan que el potencial que brinda la tecnología en las clases va a depender no solo de la naturaleza del dispositivo sino también de la gestión de la clase es decir como es relacionada con otros medios y de la cultura de la clase en la que los docentes, estudiantes y los dispositivos interactúan para producir y adquirir el conocimiento.

Beneficios de la tecnología en la enseñanza de la Estadística.

Blanco (2018) menciona que gracias a la tecnología el modo en que se ha llevado a cabo la enseñanza de la Estadística ha cambiado drásticamente, un ejemplo de ello es la liberación de tiempo dedicado a las rutinas de cálculo y la posibilidad de trabajar con grandes bancos de datos. Este autor insta a un uso adecuado de la tecnología que ayude principalmente al aprendizaje y al modo de pensar y comprender información estadística. Siguiendo con esta

misma idea los autores Cuétara et al. (2016) hacen referencia a que con el desarrollo de la informática y de los llamados paquetes estadísticos se ha generado una simplificación del trabajo de procesamiento estadístico, puesto que estos son capaces de hacer en cortos periodos de tiempo grandes operaciones matemáticas y simplemente dejar al hombre la interpretación de los resultados.

Así mismo, Rodríguez et al. (2016) considera que la tecnología dentro de la estadística permite una mejor comunicación e interacción con la sociedad, fomenta la calidad del aprendizaje y del desarrollo de destrezas, hasta incluso permitiendo el desarrollo de nuevos modelos pedagógicos basados en el uso de las capacidades y potencialidades que ofrece la tecnología.

En sintonía Gea et al. (2015) destacan que la tecnología dentro de la enseñanza de la estadística brinda una variedad de beneficios, entre ellos: permiten resolver problemas más complejos y en contexto real; fomentan en el estudiante el uso del lenguaje (verbal, simbólico, icónico, gráfico); facilitan y amplían la gama de procedimientos a su alcance; hacen ostensivos conceptos y propiedades, sobre todo por medio de la simulación y promueven el razonamiento y la argumentación.

Así mismo para Medina et al. (2013) la tecnología actual ha permitido llevar los conocimientos a los estudiantes a cualquier hora y en cualquier lugar, puesto que tienen acceso a una amplia variedad de ejercicios interactivos y cuando ellos lo requieran pueden estar recibiendo explicaciones sobre temas en específicos. Al tener a su disposición recursos didácticos que usan tecnología, los estudiantes van a valorar mucho tales recursos y a mejorar su proceso de enseñanza y aprendizaje.

Alpízar (2007) menciona que entre las herramientas tecnológicas existen una amplia variedad destinadas a apoyar la enseñanza de la Estadística, entre ellas: paquetes estadísticos, tutoriales y recursos en internet. Según este autor, los paquetes estadísticos son conocidos como herramientas, aquí se incluyen lo que son softwares dinámicos, que suelen ser utilizados para calcular medidas estadísticas, construir diversas representaciones y simular poblaciones, y generalmente cuentan con el formato de una hoja de cálculo en donde se introducen los datos.

Softwares en la enseñanza de la Estadística.

El software es definido por Sommerville (2007, como se citó en Marcano y Benigni, 2014) como aquel conjunto de programas de computadora que son desarrollados en atención a las necesidades de los clientes, a su vez pueden estar dirigidos a un mercado en general. Además de ello, Friel (2007), sostiene que el software debe ser una herramienta que facilite la

incorporación de ideas y conceptos de la disciplina. Así, el tiempo que el alumno ahorra al no tener que embarcarse en faenas de cálculos, debería reinvertirse en la exploración y práctica de conceptos estadístico.

Rouquette y Suárez (2018) reconocen que uno de los recursos tecnológicos que mayores beneficios aporta al ser utilizado en la enseñanza de la Estadística son los softwares estadísticos. Si bien es cierto este tipo de programas no fueron creados con fines pedagógicos como los softwares educativos, su uso puede potencializar al máximo el aprendizaje de contenidos estadísticos, y a la par proporcionar herramientas que muy probablemente llegará a usar el estudiante una vez que se integre al mercado laboral. Asimismo, mencionan que la diversidad de programas que existen actualmente facilita su incorporación tanto en las aulas de clase como en el campo profesional, debido a que cada uno de estos softwares poseen ventajas y propiedades específicas, además de que pueden ser elegidos de acuerdo a los intereses que tengan en particular.

Según Romeu (2013) el hacer uso de algún software ya sea en particular, estadístico o de simulación permite resolver dos problemas, por un lado, ofrece al profesor un mayor número de ejemplos prácticos, debido a que no tiene que resolverlos a mano y por el otro permite al alumno concentrarse solamente en el razonamiento estadístico.

R, Excel, SAS, Python y demás, son programas de softwares diseñados con el fin de ejecutar análisis estadísticos y generar gráficos que se pueden utilizar para diversos fines ya sean académicos, investigativos, entre otros. A continuación, en la tabla 1 se detallan y clasifican cada uno de estos Softwares:

Tabla 1

Softwares estadísticos más conocidos.

| Categoría | Software | Descripción | Aplicación |
|----------------------------|-----------------|--|--|
| Interfaz de usuario | <u>SPSS</u> | Trujillo et al. (2015) menciona que SPSS (Statistical Package for the Social Sciences) es un paquete estadístico que engloba una amplia variedad de programas capaces de realizar desde un simple análisis descriptivo hasta diferentes tipos de análisis multivariantes de datos, además de ello permite obtener tablas de frecuencias. | Lázaro et al. (2021) llevo a cabo un estudio en el cual hizo uso del software SPSS para identificar factores predictivos de deserción estudiantil. Utilizando métodos como histórico lógico, análisis síntesis y por supuesto la Estadística Descriptiva e Inferencial. Tras haber hecho uso del software, vieron la gran utilidad que este aporta puesto que, consideraron que la |

| Categoría | Software | Descripción | Aplicación |
|-----------|---------------|---|---|
| | | <p>En un principio SPSS fue diseñado para las ciencias sociales, pero con el pasar del tiempo se observó que su aplicación se extendía a la mayoría de las ramas de la ciencia.</p> <p>Cabe mencionar que este Software no es gratuito.</p> | <p>tecnología como herramienta para analizar datos del fenómeno científico en estudio relacionado con aspectos sociales, permitió entender y aportar nuevos conocimientos.</p> |
| | <u>JASP</u> | <p>Según Goss y Meneses (2019) JASP se trata de un paquete estadístico de código abierto multiplataforma, el cual fue desarrollado y actualizado ininterrumpidamente por un grupo de investigadores de la Universidad de Amsterdam. Su objetivo principal erandesarrollar un programa libre y de código abierto que incluyera tanto los estándares como las técnicas estadísticas más avanzadas, poniendo especial énfasis en desarrollar un interfaz de usuario simple e intuitiva.</p> | <p>Cahyono et al. (2021) consideró necesario realizar una formación sobre el análisis de datos de investigación utilizando los programas JASP y SPSS para que pueda ayudar a que a los estudiantes les resulte más fácil analizar los datos. Para ello, llevo a cabo capacitaciones y tutorías para estudiantes con el fin de mejorar sus habilidades para realizar investigaciones y procesar datos estadísticos. Los resultados mostraron que el 88,9 % de los estudiantes fueron capaces de comprender y analizar bien las funciones de los programas JASP y SPSS.</p> |
| | <u>Jamovi</u> | <p>Santabábara y Lasheras (2020) mencionan que Jamovi es un software de acceso gratuito y multiplataforma que utiliza el lenguaje de programación de R, además proporciona una interfaz amable con el usuario, similar a la de SPSS, afianzándose como una alternativa potente a éste. Adicionalmente, ofrece la posibilidad de visualizar la sintaxis, sirviendo también como puente para usuarios familiarizados con interfaces de menús que deseen iniciarse en el lenguaje de comandos de R</p> | <p>Dogan y Can (2019) llevaron a cabo un informe en el cual presentaron las características fundamentales del software gratuito Jamovi a los académicos en el campo de la medición educativa para su uso en investigaciones a nivel tanto de pregrado como posgrado. Entre las características más destacadas tenemos: proporciona funciones básicas como entrada y manipulación de datos, filtrado de datos basado en reglas y transformación de variables; compatible con formatos de archivos de datos populares</p> |

| Categoría | Software | Descripción | Aplicación |
|---------------------|---------------|---|--|
| | | | como csv, RData, dta y sav; es capaz de realizar muchos análisis de variables únicas y múltiples. |
| Codificación | <u>R</u> | Según Avello y Seisdedo (2017), R (<i>R Development Core Team</i>) es un programa estadístico y un lenguaje de programación de uso libre, gratuito y código abierto, el cual fue desarrollado a partir de un proyecto colaborativo voluntario de investigadores y estadísticos de diversos países. Es un programa basado en comandos, que permite acceder a todos los procedimientos y opciones a través de una sintaxis textual. Fue oficialmente presentado en 1997 bajo Licencia General Pública de la Fundación de Software Libre | López (2016) llevo a cabo un estudio con el uso del Software R para enseñar Estadística. Este autor menciona que R posee un increíble potencial para ser usado en las aulas de clase, es una herramienta útil debido a su capacidad para hacer cálculos estadísticos, crear gráficos y sobre todo, por la posibilidad de trabajar desde la estadística elemental hasta las estadísticas más avanzadas. Los resultados que arrojaron este estudio fueron que los participantes reconocieron en el programa su utilidad como herramienta y, cuando fueron llevados a utilizarlo realizaron comentarios muy favorables sobre sus potencialidades didácticas |
| | <u>SAS</u> | Rodríguez Y Díaz (2009) mencionan que el software de SAS, es un sistema de entrega de información que provee acceso transparente a cualquier fuente de datos, incluyendo archivos planos, archivos jerárquicos, y los más importantes manejadores de bases de datos relacionales. Además de esto, también incluye su propia "data warehouse", es decir una base para almacenar y manejar los datos. | Rodríguez Y Díaz (2009) en su estudio sobre la minería de datos consideran que el Software SAS es de gran ayuda, puesto que el sistema soporta un amplio rango de aplicaciones, destacándose el análisis estadístico, análisis gráfico de datos, análisis de datos guiado, mejoramiento de la calidad, diseño experimental, administración de proyectos, programación lineal y no lineal, generación de reportes, aplicaciones de multimedia, entre otros. |
| | <u>Python</u> | Para Marzal et al. (2014) Python un lenguaje de programación de muy alto nivel, muy expresivo, legible, | Alayo et al. (2021) en su estudio menciona que debido al carácter especializado de algunas herramientas |

| Categoría | Software | Descripción | Aplicación |
|--|--------------|--|--|
| | | <p>sintaxis muy elegante y permite la escritura de programas cuya lectura resulta más fácil que si utilizáramos otros lenguajes de programación, además de ello, ofrece un entorno interactivo que facilita la realización de pruebas y ayuda a despejar dudas acerca de ciertas características del lenguaje, su entorno de ejecución detecta muchos de los errores de programación y proporciona información para detectarlos y corregirlos.</p> | <p>tecnológicas y otros inconvenientes que hacen difícil la actualización de los contenidos y la motivación de los estudiantes, propone el uso de Python para la enseñanza de las matemáticas y materias afines. Este autor reconoce que Python se ha convertido en uno de los lenguajes más utilizados en la actualidad debido a su versatilidad. Es fácil de aprender y cuenta con una comunidad muy activa que ha provisto herramientas potentes y útiles en varios dominios de aplicación</p> |
| Mixto (Codificación e interfaz) | <u>Excel</u> | <p>Según Alfaro y Alfaro (2018) Excel es un programa informático desarrollado y distribuido por Microsoft Corp. que permite realizar tareas contables y financieras mediante hojas de cálculo, forma parte de Microsoft Office y es utilizado en gran parte del mundo. La accesibilidad y flexibilidad que posee Excel constituye una herramienta complementaria para el análisis de los conceptos y materias estadísticas.</p> | <p>López et al. (2015) en su estudio sobre el uso de herramientas computacionales considera que Excel es sin duda un excelente programa para llevar a cabo la enseñanza de la Estadística, puesto que con la implementación de este programa hace que los estudiantes dejen de considerar a esta disciplina como difícil de aprender, gracias a esta pueden aplicar los conceptos estadísticos que se les imparten en el trabajo investigativo que realizan, en diversas situaciones del mundo real.</p> |
| | <u>Stata</u> | <p>StataCorp (2023) menciona que el software estadístico stata provee todas las herramientas para la gestión, análisis y visualización de datos asociadas a una interfaz gráfica potente y a la vez fácil de usar. Además de ello, cuenta con una amplia variedad de procedimientos estadísticos que pueden ser llevados a cabo en diferentes</p> | <p>Este Software ha sido utilizado en varios estudios como por ejemplo en Améstica et al. (2014) y Laureano (2021) quienes hicieron uso de este software para trabajar con bases de datos y poder hacer estimaciones en base a los resultados. Como conclusiones ambos estudios concuerdan que el uso de las TIC mejora significativamente el nivel de</p> |

| Categoría | Software | Descripción | Aplicación |
|-----------|----------|---|--|
| | | áreas y sectores, por lo que es muy utilizado no solo en investigación académica, sino también en entidades gubernamentales, financieras, comerciales y hasta de servicios. | comprensión de conceptos en los alumnos. |

En la enseñanza de la Estadística, Medina et al. (2022) menciona que tres de los recursos o medios didácticos más importantes para la formación de conceptos son: los juegos, el uso de herramientas computacionales y los datos. Este último recurso puede considerarse como el corazón del trabajo estadístico puesto que, Estrella (2017) considera que los sujetos necesitan comprender que los datos son números con un contexto, además de ello reconocer la necesidad de datos para tomar decisiones y evaluar la información. Este autor menciona que los diferentes tipos de datos, los métodos para recolectarlos y producirlos, hacen la diferencia en los tipos de conclusiones que se pueden.

Uso de datos abiertos.

Según Wilson y Cong (2021) los datos abiertos, conocido por sus siglas en inglés OGD Open Government Data, constituyen un entorno relacionado con el marco legal, en el uso de tecnologías de información y la participación de las partes interesadas como la ciudadanía, la empresa e incluso las instituciones públicas.

D'Agostino et al. (2017) destaca algunos aspectos positivos de los datos abiertos, como son: reducción de brechas de información y fortalecimiento de vínculos entre las instituciones de gobierno y los representantes de la sociedad civil; fortalecimiento de los procesos de investigación y traducción del conocimiento tanto en instituciones académicas, científicas como gubernamentales; formulación de políticas públicas basadas en evidencias; y, desarrollo de campañas de comunicación social más efectivas con mejor uso de las tecnologías de información y redes sociales.

Agregado a lo anterior Zuiderwijk et al. (2015) menciona que los OGD se puede utilizar para cinco actividades principales, como son: para buscar y encontrar OGD (navegar, consultar y explorar conjuntos de datos); analizar OGD (análisis estadístico, transformación de datos, visualización de datos en línea y descarga de datos); visualizar OGD (generación de diagramas, mapas, gráficos, representaciones interactivas de conjuntos de datos); interactuar con los OGD (colaboración a través de foros de discusión, mensajes, grupos de usuarios, entre otras

funcionalidades) y analizar la calidad de OGD (análisis y evaluación de la calidad del conjunto de datos).

Según Thorsby et al. (2017) los datos abiertos deben cumplir con ciertos requisitos para ser utilizados de forma correcta, entre ellos: que los datos deben estar completos, ser información primaria, ser accesibles desde internet, poseer la capacidad de ser procesados en un ordenador, no se debe discriminar su acceso, no deben poseer un propietario y licencia para su acceso.

Muente y Serale (2018) consideran que los datos abiertos son esenciales para mejorar la confianza de los ciudadanos en sus estados y hacerles partícipes de la acción de decisiones públicas, facilitando así el acceso a la información y mejorando la calidad de los servicios públicos. Actualmente, la creciente apertura de los datos abiertos ha construido una nueva concepción del Estado centrado en los ciudadanos. En tal sentido, los gobiernos deben asegurar la calidad y confiabilidad de los datos y crear espacios de colaboración en donde se tome en cuenta el valor del público.

Los conjuntos de datos abiertos tienen un gran potencial según Marín (2022), ya que suelen ser utilizados para desarrollar recursos educativos abiertos (REA), los cuales son materiales de enseñanza, aprendizaje e investigación en soporte digital cuyo carácter es gratuito, pues son publicados bajo una licencia abierta que permite su uso, adaptación y redistribución para usos no comerciales de acuerdo a la UNESCO.

Datos abiertos Ecuador.

La plataforma “Datos abiertos Ecuador” es una iniciativa de Gobierno Abierto que surge de la necesidad de contar con un rol más activo y participativo de la sociedad en la toma de decisiones y solución de problemas públicos, así como de una gestión de puertas abiertas por parte del Estado y la generación de oportunidades de participación, esto con el fin de que los gobiernos sean más transparentes, responsables y eficaces, lo que conlleva a un nuevo modelo de gobierno en el que se promueve la participación de la ciudadanía y una administración pública más cercana a las personas. Actualmente en el Ecuador existen 118 entidades públicas, de las cuales 95 liberan sus bases de datos en la plataforma de Datos abiertos Ecuador, gracias a la colaboración de estas entidades se permite el aporte al fortalecimiento de la transparencia y acceso a la información, integridad y rendición de cuentas, colaboración e innovación y participación ciudadana.

Los datos abiertos han sido destacados como clave para la transparencia de la información y el avance científico según Atenas y Havemann (2015), puesto que los

estudiantes al estar expuestos al uso de los Datos Abiertos tienen el acceso a las mismas materias primas que son utilizadas por científicos y responsables políticos, lo que les permite enfrentar problemas reales tanto a nivel local como global. Con seguridad este autor reconoce que, cuando los educadores utilizan los datos abiertos en la enseñanza y aprendizaje motivan a los estudiantes a tener un pensamiento como investigadores, periodistas, científicos y como responsables políticos y activistas. Además de ello, proporcionan un contexto significativo para poder adquirir experiencia en trabajos y procesos investigativos, así como para aprender buenas prácticas en gestión, análisis y presentación de informes de datos.

Vila (2018) menciona que el proyecto “School of Data”, relacionado con el conocimiento abierto es una iniciativa la cual además de brindar orientación y apoyo a los estudiantes, educadores, investigadores, entre otros, que trabajan con datos abiertos en contextos educativos, ofrece cursos en línea con el fin de comprender los datos abiertos. Además, ofrece recursos que han sido traducidos al español, francés y portugués, con el fin de ayudar a un número significativo de educadores a nivel internacional.

Implementación de datos abiertos en la enseñanza de la Estadística con Python.

De acuerdo con Hall (2011) existen muchos beneficios y desafíos que son exclusivos de la enseñanza y el aprendizaje de Estadística con datos reales, incluida la participación de los estudiantes y el conocimiento adquirido por los mismos cuando trabajan con datos reales. Los datos que son recopilados y analizados por los estudiantes, adquieren un campo completamente nuevo de interés y relevancia. Sin embargo, no todos los datos reales tienen el mismo atractivo para los estudiantes, por ello es esencial utilizar datos que sean de interés para estos. Además de ello, el autor menciona que, al trabajar con datos reales, se facilita su comprensión y se logran habilidades estadísticas mejorando habilidades estadísticas enormemente y más aún cuando participan en las fases iniciales de planificación de algún proyecto.

Asimismo, Marín (2022) considera que los datos abiertos cuando son utilizados como recurso educativo permite que los estudiantes aprendan y experimenten trabajando con los mismos conjuntos de datos que emplean investigadores, gobiernos y sociedad civil. Es decir, al trabajar con datos reales, los estudiantes desarrollan y aplican habilidades analíticas y colaborativas para resolver problemas del mundo real.

En el trabajo titulado “Teaching Python for Data Science: Collaborative development of a modular & interactive curriculum”, de Duda et al. (2021), se utilizó un plan de estudios diseñado para estudiantes de secundaria sin experiencia previa en codificación que estén interesados en aprender Python para ciencia de datos. Tras culminar el plan de estudios se

aplicó una encuesta a los estudiantes de forma y como resultado obtuvieron que la mayoría sentía que sus habilidades tanto en programación Python, habilidades estadísticas, resolución de problemas, pensamiento crítico y colaboración habían mejorado significativamente. De hecho, un alumno compartió que planea ingresar a la universidad con el fin de estudiar Ciencias de la Computación y obtener una especialización en robótica, alrededor de un tercio de los participantes afirmaron que están más interesados en profundizar sobre la ciencia de datos.

Podworny et al. (2022) en su estudio “A place for introduction to Data Science at school: Between Statistics and Programming”, explicita la ejecución de una unidad didáctica de la ciencia de los datos centrada en el análisis de datos ambientales, que integraron en una clase de informática de noveno grado con contenidos estadísticos. Tras haber culminado la unidad didáctica, los estudiantes mencionaron que las actividades de programación en este proyecto les permitió adquirir nuevos conocimientos sobre su propio entorno. Los autores de este estudio consideran que los estudiantes de aproximadamente 14 años de edad pueden trabajar con datos utilizando una herramienta de programación de nivel profesional, siempre que su entorno esté configurado adecuadamente. Con base en los resultados concluyen que las clases de informática pueden ser un lugar para introducir la ciencia de datos, incorporando elementos de estadística y programación.

En el estudio “Integrated Data Science for Secondary Schools: Design and Assessment of a Curriculum” realizado por Schanzer et al. (2022) llevaron a cabo un plan de estudios de ciencias de datos en la escuela secundaria el cual se basó en cuatro elementos esenciales; programación, Estadística, dominios significativos y responsabilidad cívica. Los dos primeros cubren habilidades técnicas básicas; el tercero es importante para motivar a los estudiantes, permitir su uso en múltiples contextos y tener impacto; y el cuarto es necesario tanto para comprender las aplicaciones como para evitar los peligros de utilizar datos sin tener en cuenta sus impactos sociales. Los resultados de estudio hacen referencia a que tanto los profesores como alumnos a menudo han encontrado mayor autenticidad y entusiasmo por la programación, de hecho, en la mayoría de los casos, los estudiantes en lugar de elegir los conjuntos de datos que constan en el plan de estudios, algunos de ellos descargaron datos nuevos de la Web, datos que son de interés para los mismos y con los cuales pueden aprender de mejor forma.

Fleischer et al. (2022) en su estudio “Teaching and learning data driven machine learning with educationally designed Jupyter notebooks”, explicita la ejecución de un curso de ciencia de datos en un curso de educación secundaria, el estudio exploró como los estudiantes aplicaron un aprendizaje automático haciendo uso de Jupyter Notebook, documentando el

proceso en un ensayo computacional. Durante el curso los estudiantes primero aprendieron los conceptos básicos de Python y habilidades en el uso de CODAP, además examinaron conjuntos de datos que planteaban preguntas estadísticas y al final trabajaron en un proyecto en el que debían elaborar un ensayo. Como resultados obtuvieron que en la mayoría de los ensayos todos los estudiantes pudieron establecer una buena narrativa para su proceso de aprendizaje automático.

En el estudio “Experimental Functionality Development for Scratch Mathematical and Statistics Extensions” realizado por Chang et al. (2016), llevaron a cabo la implementación de Scratch y Python para la comprensión de contenidos matemáticos y estadísticos. Tal estudio se centró principalmente en el desarrollo de las habilidades de pensamiento de los estudiantes en computación y conceptos lógicos. Además, según los autores la combinación de la tecnología y educación, brinda a los estudiantes una amplia variedad de habilidades tanto en programación como de pensamiento computacional.

En el trabajo titulado “Data science education in secondary schools: Teaching and learning decision trees with CODAP and jupyter notebooks as an example of integrating machine learning into statistics education”, de Biehler et al. (2020), se explicita la ejecución de un curso de ciencia de datos el cual constaba de tres módulos. El primer módulo trataba sobre los conceptos básicos de estadística y exploración de datos, el segundo sobre el aprendizaje automático y el tercero consistía en un trabajo por proyectos en el cual los alumnos aplicaban las habilidades y conocimientos previamente adquiridos a una pregunta con datos reales. Tras finalizar el curso los estudiantes quedaron aptos para evaluar estadísticamente un modelo de árbol de decisión y reflexionar sobre las implicaciones sociales que pueden tener.

Lasser et al. (2023) en su estudio “Introductory data science across disciplines, using Python, case studies, and industry consulting projects”, se llevó a cabo la ejecución de un curso de ciencia de datos cuyo objetivo fue proporcionar a todos los estudiantes una comprensión fundamental de los conceptos y procedimientos de la ciencia de datos y motivarlos a seguir algunos cursos adicionales orientados a desarrollar sus competencias en este ámbito. En tal estudio consideraron el uso de Python ya que es un lenguaje de programación sintácticamente simple, el cual facilita el aprendizaje de conceptos básicos de programación. Tras finalizar el curso los estudiantes reportaron niveles muy altos de interés y dijeron haber aprendido mucho durante las clases vistas. En las conversaciones personales, los estudiantes se mostraron especialmente positivos con respecto a la posibilidad de algún día poder trabajar con empresas y grupos de investigación.

Vinces y Ríos (2022) en su estudio “Cuadernos interactivos: experiencia de enseñanza y aprendizaje de estadística en el entorno Jupyter”, se llevó a cabo una experiencia de enseñanza y aprendizaje de estadística descriptiva en el entorno Jupyter Notebook, en la cual se usó una base de datos abiertos sobre el registro de vehículos en el servicio de rentas internas. Entre los resultados se evidencia que los estudiantes encontraron una forma nueva y efectiva de aprender, permitiéndoles comprender conceptos como el manejo de bases de datos y el uso de técnicas visuales, además de una mayor predisposición de los estudiantes a adentrarse en la asignatura.

Tang (2020) en su estudio “A simple introduction to programming and statistics with decision trees in R”, llevó a cabo una actividad docente de 45 minutos de duración, la cual otorgaba a los estudiantes conocimientos prácticos sobre estadística y lenguaje de programación. Los materiales didácticos empleados se basaban en el lenguaje de programación R y Python, este trabajo llegó a ser considerado beneficioso tanto para docentes como estudiantes, por un lado los estudiantes pueden explorar conceptos estadísticos y de programación, si los mismo lo llegan a considerar útil y relevante pueden incorporarse en la enseñanza de la misma. Por otro lado, los docentes pueden observar la reacción que tienen los alumnos y evaluar con el fin de que se sugiriera la implementación de estos recursos en la educación a fin de mejorar la experiencia de aprendizaje de esta disciplina.

En el trabajo titulado “Use of commercial and free software for teaching statistics”, de Stemock y Kernes (2019) se realizó con fin de demostrar que el uso de softwares estadísticos puede favorecer a la enseñanza y aprendizaje de la estadística haciendo de ella mas efectiva, generando aprendizajes activos y significativos, mejorando la comprensión de los conceptos estadísticos de los estudiantes, además de sus habilidades en la resolución de problemas. Con la realización de este estudio queda en evidencia de que a medida que surgen más softwares estadísticos se vuelve muy importante realizar investigaciones que aborden específicamente estas cuestiones esenciales relacionadas con el uso de la tecnología.

En un esfuerzo por hacer que las estadísticas sean lo más relevantes posible, Holman y Hacherl (2023) llevaron a cabo el estudio “Teaching Monte Carlo Simulation with Python”, en el cual comenzaron a enseñar un módulo de computación estadística utilizando Python, El curso se impartió durante 80 minutos dos veces por semana en un formato tradicional, como resultado obtuvieron que los estudiantes brindaron comentarios positivos entre ellos, la mejora de habilidades para la resolución de problemas y la gran utilidad de esta herramienta al momento de hacer uso de ella en situaciones de la vida real.

Estadística en el sistema educativo ecuatoriano.

En Ecuador la educación está guiada por el Sistema Nacional de Educación, el cual contempla al Sistema Intercultural Bilingüe, y el Sistema de Educación Superior. El Sistema Nacional de Educación se encuentra conformado por los niveles de educación: inicial, general básica, bachillerato y educación superior. Con el pasar del tiempo, la educación en el Ecuador se ha convertido en una educación más organizada y sistematizada todo esto es gracias a las leyes enfocadas en la educación que han sido implementadas. (Suasnabas y Juárez, 2020)

Según Mineduc (2016) en el Ecuador, la enseñanza de estadística proviene desde el año 1982 en bachilleratos técnicos de contabilidad e informática mediante materiales bibliográficos y usando el método deductivo. Posteriormente a ello, en el año 2008 se incorpora de forma obligatoria en toda la educación secundaria como el sexto bloque curricular de estadística y probabilidad, el cual se mantiene vigente hasta la actualidad con un enfoque educativo constructivista. El Ministerio de educación mediante el currículo ecuatoriano, establece la enseñanza de la matemática en todos los años de educación secundaria de forma amplia y apegada al constructivismo, en donde, se busca que el estudiante sea el principal actor en promover sus conocimientos, fortaleciendo la capacidad de razonar, analizar, sistematizar, crear y resolver problemas, esto con el fin de formar individuos autónomos, creativos, y gestores de nuevas ideas.

De acuerdo con Aguilar et al. (2021) la estadística como asignatura tiene como fin promover el beneficio humano, la calidad de vida y el bienestar mediante el uso efectivo del conocimiento derivado del manejo de datos. Esta es aplicada en los planes de estudio de las más diversas profesiones y forma parte de una variedad de programas de especialización y posgrado en las diversas disciplinas, es de gran importancia en el desarrollo de la ciencia, las tecnologías y en las más diversas esferas de la vida cotidiana.

Por ello, Álvarez et al. (2022), mencionan que en el campo específico de la estadística y probabilidad la importancia de su enseñanza en el Ecuador radica en la formación de niños, jóvenes y adultos con suficientes habilidades para leer e interpretar gráficos, tablas presentes en diferentes medios informativos sobre temas ya sean sociales, económicos o políticos, a esto se suma, la exigencia de la mayoría de profesiones que día a día buscan que sus estudiantes tengan competencias básicas. Sin embargo, dichos autores agregan que los resultados de aprendizaje de los últimos 5 años correspondientes a los ejes temáticos de estadística y probabilidad como parte del examen Ser Bachiller, han reflejado una media de 440 puntos, muy debajo del puntaje base, cuya nota mínima es de 700 o el 70% de la evaluación requerida para aprobar el bachillerato, a pesar de que han transcurrido estos

años, la asignatura ha tenido poca mejora, aumentando del 37% al 54%, esto no es suficiente ya que significa una falta de conocimientos y desarrollo de destrezas cognitivas en los estudiantes, por lo que esto evidencia posibles dificultades en la aplicación de medios didácticos en el proceso enseñanza y aprendizaje.

5. Metodología

La investigación se enmarcó en un enfoque mixto, la parte cualitativa se utilizó para extraer los resultados de las diferentes investigaciones enmarcadas en procesos de enseñanza de la Estadística Descriptiva con datos abiertos y Python, mientras que en lo cuantitativo se utilizó para cuantificar mediante análisis de frecuencias la satisfacción de un grupo de estudiantes que fueron parte de una experiencia áulica recibiendo clases de Estadística con datos abiertos procesados con Python. Según Otero (2017) “El proceso de investigación mixto implica una recolección, análisis e interpretación de datos cualitativos y cuantitativos que el investigador haya considerado necesarios para su estudio” (p. 19).

En la misma se aplica el método deductivo-inductivo, deductivo para la revisión documental sistemática sobre el uso de datos abiertos y Python en la enseñanza de la Estadística Descriptiva y el inductivo fue utilizado para recopilar datos y observar patrones en el proceso de enseñanza de la estadística utilizando datos abiertos gestionados con Python. Además, el alcance de la investigación es de tipo exploratorio-descriptivo. Exploratorio porque no se han llevado a cabo estudios previos sobre la enseñanza de la Estadística con datos abiertos mediante Python en el nivel secundario de educación. Descriptivo ya que se siguió un proceso de revisión documental sistemática, mediante el uso de técnicas de organización de información como la bitácora de búsqueda, fichas bibliográficas y de contenido, teniendo como unidades de análisis los documentos científicos. La naturaleza de este trabajo se enmarca en un tipo no experimental, dado que no se manipula las variables que constituyen el objeto de estudio.

La población estuvo compuesta 25 estudiantes que cursaban el 2do año de Bachillerato General Unificado (BGU) de un colegio público del Sur de Ecuador. Este grupo de estudiantes se eligieron por conveniencia, bajo el criterio que recibían un mayor número de horas clase, adicionalmente, este grupo aún no recibía los contenidos de Estadística Descriptiva por lo que las autoridades condicionaron a que se imparta las experiencias áulicas a este grupo; de tiempo, considerando que para entregar el informe de esta investigación tan solo se tiene 16 semanas y, de pertinencia, es decir, que los estudiantes sean del segundo año de BGU. Previo a elegir al grupo se tuvo que establecer un convenio entre la Universidad y la Institución Educativa, realizando en primera instancia un acercamiento con la máxima autoridad de la institución a fin de tener un diálogo verbal para luego formalizarlo a través de un escrito (Anexo2).

Para dar cumplimiento al primer objetivo específico, que consistió en indagar los diferentes procesos de enseñanza de la Estadística Descriptiva en nivel secundario utilizando datos abiertos gestionados con Python, se realizó una revisión documental sistemática recolectando fuentes de información como: tesis, artículos de libro, revistas científicas , entre

otros, que estén relacionadas con el tema de estudio, las cuales se revisaron detalladamente para seleccionar las que mejor se acople a las necesidades y características de la investigación. La búsqueda de información se la realizó en motores de búsqueda y bases de datos científicas como: Scopus, Scielo, Redalyc, Google Académico, usando las siguientes ecuaciones de búsqueda avanzada, tales como las que se muestran en la Tabla 2:

Tabla 2

Ecuaciones de búsqueda utilizadas

| Nº Ecuación | Ecuación de búsqueda | Nº documentos seleccionados |
|-------------|--|-----------------------------|
| 1 | Python AND statistics AND secondary AND education | 4 |
| 2 | (TITLE-ABS-KEY (teaching of statistics) AND TITLE-ABS-KEY (with Python ion secondary education)) | 2 |
| 3 | Teaching of statistics AND Python AND Data Science AND secondary education | 5 |

Cabe indicar que, para la organización de la información se utilizó una bitácora, con la siguiente estructura: Motor de búsqueda, Ecuación de búsqueda, N. documentos, N. documentos seleccionados, Tipo de documento, Año, Autor/es, Título y doi o url (Anexo 3). Para la selección de los documentos se tomó en cuenta los siguientes criterios: relevancia del tema, calidad del documento, accesibilidad del documento, claridad de la presentación y posición en los resultados de búsqueda. De tal manera que de la ecuación 1 se seleccionaron 4 documentos, de la ecuación 2 se seleccionó 2 y de la ecuación 3 se seleccionó 5, dándonos un total de 11 documentos. Los criterios de exclusión fueron: investigaciones con metodologías no claras, las palabras claves no tenían nada que ver con el estudio, resultados no claros. Con el análisis de datos se realizó una síntesis de los resultados obtenidos (Figura 3), resaltando los aspectos más relevantes y conclusiones esenciales de la revisión documental sistemática, esto ayudó a garantizar la calidez de los resultados, apoyando de esta manera a la propuesta de solución que se llevó a cabo en las siguientes etapas de investigación.

Para dar alcance al segundo objetivo específico que se refiere a identificar el grado de satisfacción de los estudiantes cuando aprenden estadística con datos abiertos y Python se realizó una experiencia áulica de 6 clases teóricas-prácticas con los temas estadísticos de: *medidas de tendencia central, medidas de dispersión y técnicas visuales de datos*, donde se siguió un proceso de enseñanza sugerido por la revisión documental de la fase 1, es decir que se tomaron las investigaciones en la que se hayan enseñado estadística con Python y datos abiertos y que obtuvieron resultados favorables en el aprendizaje.

Durante las 3 primeras clases se abordó los temas mencionados de manera magistral, a partir de la clase 4 se habló acerca de un conjunto de datos llamados datos abiertos, presentando situaciones en la que los estudiantes llegarán a hacer uso de los mismos, sin embargo, al ser demasiado grandes se genera la necesidad de utilizar alguna herramienta para procesarlos dado que ya no se los puede manejar mediante una hoja física de cálculo e incluso siendo pesados para herramientas como Excel. Para abordar este caso, se propuso utilizar el software Python desde un enfoque educativo para dar tratamiento a estas bases de datos voluminosas de manera eficiente y a su vez se procedió a realizar una demostración práctica utilizando el conjunto de datos “Estadísticas Vehículos 2022”. En la clase 5, los estudiantes tuvieron la oportunidad de utilizar y aplicar Python para la resolución ejercicios estadísticos con ayuda del docente (en este caso, el autor de esta investigación), a la par, mediante una interacción docente-alumno se exploró las características de los datos de “inocar_altura_horaria_predicha_2023diciembre”, empleando la herramienta para llevar a la práctica los conocimientos de medidas de tendencia central, medidas de dispersión y técnicas visuales de datos.

Al final de esta experiencia áulica, en la clase 6 se abrió un espacio en la que los estudiantes compartieron sus opiniones sobre el uso de Python y los datos abiertos durante las clases dadas y luego se aplicó un instrumento para determinar el grado de satisfacción. Por cuestiones de tiempo tanto de los estudiantes como de quien hizo este trabajo no se pudo aplicar un test.

Para determinar el grado de satisfacción se utilizó la técnica de la encuesta como medio de recolección de datos y el instrumento una escala de actitudes. Según Reyes et al. (2018), esta escala es generalmente usada como mecanismo para la recolección de información, suele ser utilizada en investigaciones para medir el grado en que se dan las categorías, actitudes, dimensiones o características. Esta escala de actitudes permitió determinar y evaluar el grado de satisfacción de los estudiantes del nivel secundario en relación con el estudio de Estadística Descriptiva utilizando datos abiertos gestionados con Python. Los datos recolectados fueron organizados y analizados mediante frecuencias lo que permitió obtener conclusiones relevantes para la investigación.

Una vez obtenidos los resultados de la investigación, se procedió a realizar una propuesta didáctica que sea conveniente para abordar el tema de enseñanza de la Estadística Descriptiva utilizando datos abiertos gestionados con Python en el nivel secundario (Anexo 1)

6. Resultados

Para la identificación sobre los diferentes procesos de enseñanza de la Estadística Descriptiva en nivel secundario utilizando datos abiertos gestionados con Python, la búsqueda y análisis de información en varias fuentes dio como resultado de 11 documentos. Para la representación de información del primer objetivo sobre los diferentes procesos de enseñanza de la Estadística Descriptiva se construyó la siguiente tabla:

Tabla 3.

Procesos de enseñanza de la Estadística Descriptiva en nivel secundario utilizando datos abiertos con Python

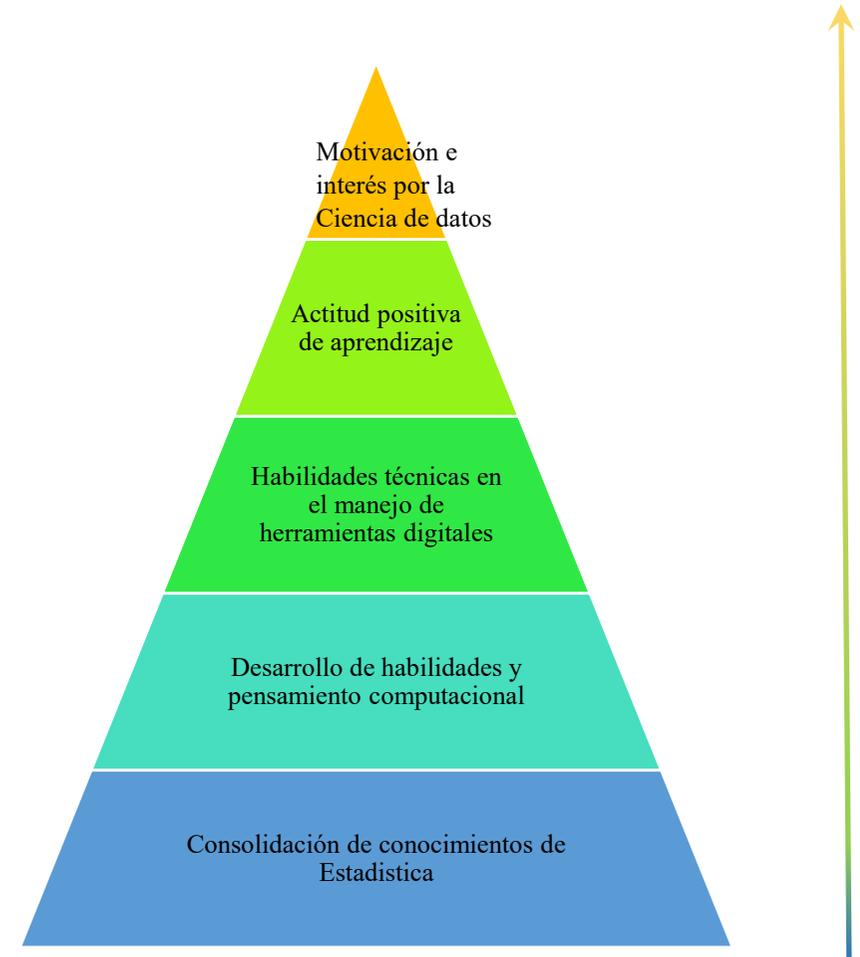
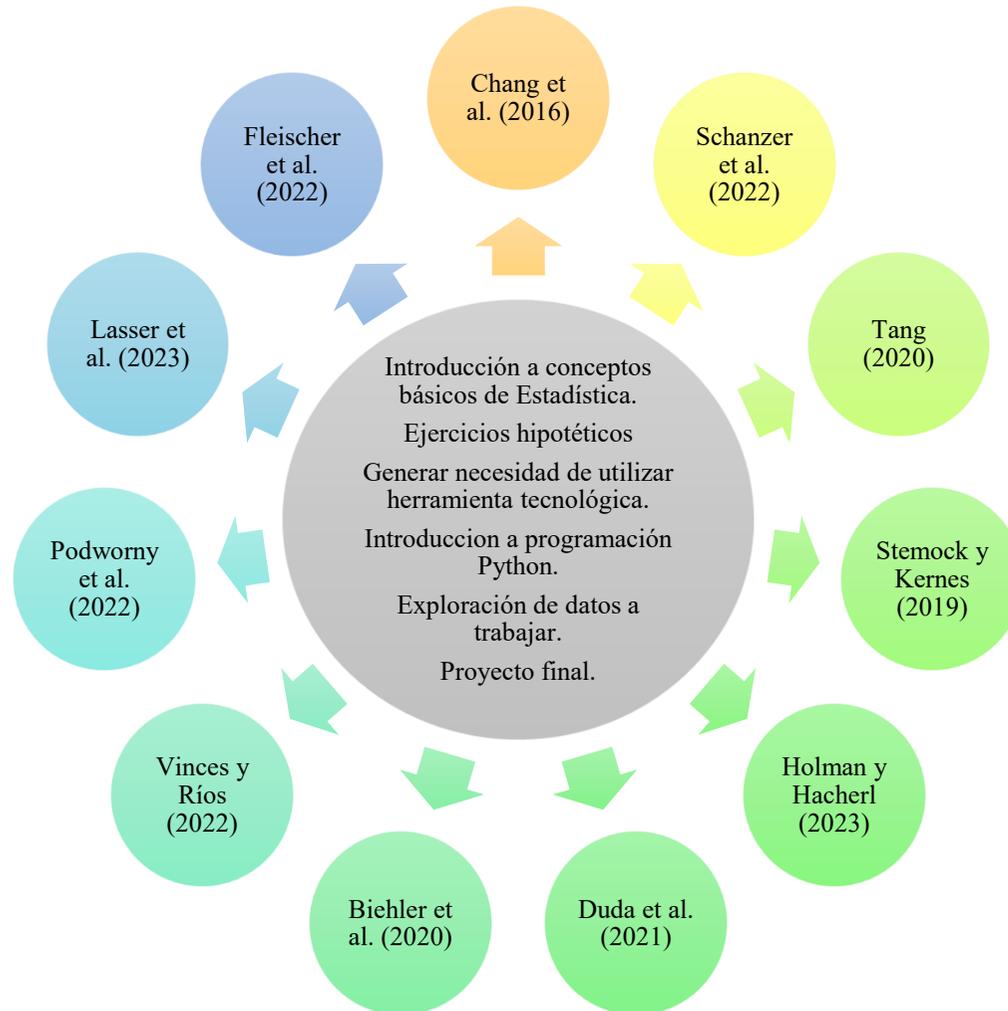
| Autor/es | Proceso de enseñanza | Tipo de investigación | Resultado de la investigación |
|---|--|------------------------------|---|
| Yannik Fleischer, Rolf Biehler y Carsten Schulte (2022) | Introducción a conceptos básicos de estadística y programación Python. Exploración de datos a trabajar. Finalmente, elaboración de proyecto final. | Investigación de campo | Habilidades técnicas en el manejo de herramientas digitales para tratamiento de datos. |
| Jana Lasser, Debsankha Manik, Alexander Silbersdorff y Benjamin Säfken (2023) | Introducción a los estudiantes con conceptos básicos de programación y estadística. Presentación de ejemplos de aplicación a la ciencia de datos. Aplicación de conocimientos en un proyecto final. | Investigación de campo | Despertaron el interés por la ciencia de datos, las clases la valoraron de alto interés para su formación académica y tuvieron una actitud positiva al momento de aprender. |
| Susanne Podworny, Sven Hüsing, Carsten Schulte (2022) | Conocimientos básicos de Estadística previamente adquiridos. Introducción a conceptos básicos de Python y Jupyter Notebooks. Exploración de datos con Python mediante Jupyter Notebooks. Elaboración de proyecto final. | Investigación de campo | Aumento de confianza y valoración en habilidades con respecto a herramientas digitales, motivación e interés por la ciencia de datos y la estadística. |
| Ching Chang, Ya-Lun Chin, Chih-Kai Chang (2016) | Enseñanza de conceptos básicos de Estadística. Introducción a Scratch y Python. | Investigación de campo | Desarrollo de habilidades y pensamiento computacionales, |

| Autor/es | Proceso de enseñanza | Tipo de investigación | Resultado de la investigación |
|---|---|---------------------------|--|
| | Desarrollo de habilidades de programación. Elaboración de proyecto final (diseño de un juego). | | razonamiento y pensamiento lógico. |
| Emmanuel Schanzer, Nancy Pfenning, Flannery Denny, Sam Dooman, Joe Gibbs Politz, Benjamin S. Lerner, Kathi Fisler y Shriram Krishnamurthi (2022) | Desarrollo de habilidades de programación y conceptos estadísticos. Introducción a CODAP y Python. Exploración con un conjunto de datos reales. Proyecto final (Ensayo con el conjunto de datos anterior) | Investigación de campo | Aumento en la motivación e interés de los estudiantes en la ciencia de datos y herramientas digitales. |
| Marlena Duda, Kelly L. Sovacool, Negar Farzaneh ¹ , Vy Kim Nguyen ¹ , Sarah E. Haynes, Hayley Falk ¹ , Katherine L. Furman ³ , Logan A. Walker, Rucheng Diao, Morgan Oneka, Audrey C. Drotos, Alana Woloshin, Gabrielle A. Dotson, April Kriebel, Lucy Meng, Stephanie N. Thiede, Zena Lapp, and Brooke N. Wolford (2021) | Introducción a conceptos básicos de programación Python y conceptos estadísticos. Trabajo práctico con una base de datos. Finalmente, proyecto final en pares. | Investigación de campo | Desarrollo de habilidades de codificación e interés en la ciencia de datos. |
| Rolf Biehler, Yannik Fleischer, Lea Budde, Daniel Frischemeier, Dietrich Gerstenberger, Susanne Podworny, Carsten Schulte (2020) | Conocimientos básicos de Estadística previamente adquiridos. Introducción a conceptos básicos de Python y Jupyter Notebook. Desarrollo de un taller práctico. Proyecto final. | Investigación de campo | Desarrollo de habilidades Estadísticas y de programación para la ciencia de datos. |
| Vinces y Ríos (2022) | Introducción magistral de los conceptos estadísticos. Resolución de ejercicios hipotéticos Generar la necesidad de utilizar una herramienta tecnológica. Introducción a Python a través de Jupyter Notebook. Aplicación con una base de datos real. | Investigación cualitativa | Aumento en la motivación e interés por la Estadística Descriptiva. |

| Autor/es | Proceso de enseñanza | Tipo de investigación | Resultado de la investigación |
|-------------------------|--|----------------------------|--|
| Tang (2020) | <p>Conocimientos básicos de Estadística adquiridos. Presentación del Software Python. Proporcionar a los estudiantes un conjunto de datos Docente-alumno exploran la base de datos. El estudiante replica el contenido en su propio ordenador. Finalmente, se evalúa el desempeño mediante un proyecto final.</p> | Investigación de campo | <p>Consideran relevante y útil el uso de herramientas de programación para el estudio de la Estadística Descriptiva.</p> |
| Stemock y Kernes (2019) | <p>Conocimientos de Estadística adquiridos. Presentación y tutorial sobre el uso del software estadístico. Incorporación de simulaciones y actividades de aprendizaje utilizando el software. Finalmente, exploración de un conjunto de datos realizando análisis estadísticos descriptivos.</p> | Investigación Experimental | <p>Aumento en la comprensión profunda de contenidos Estadísticos</p> |
| Holman y Hacherl (2023) | <p>Introducción de conceptos estadísticos. Simulación con ejercicios prácticos. Introducción a Python para la ciencia de datos. Aplicación de conocimientos en un conjunto de datos real.</p> | Investigación de campo | <p>Mayor interés en la Estadística y las herramientas de programación.</p> |

Figura 3

Proceso de enseñanza de la Estadística Descriptiva con datos abiertos y Python



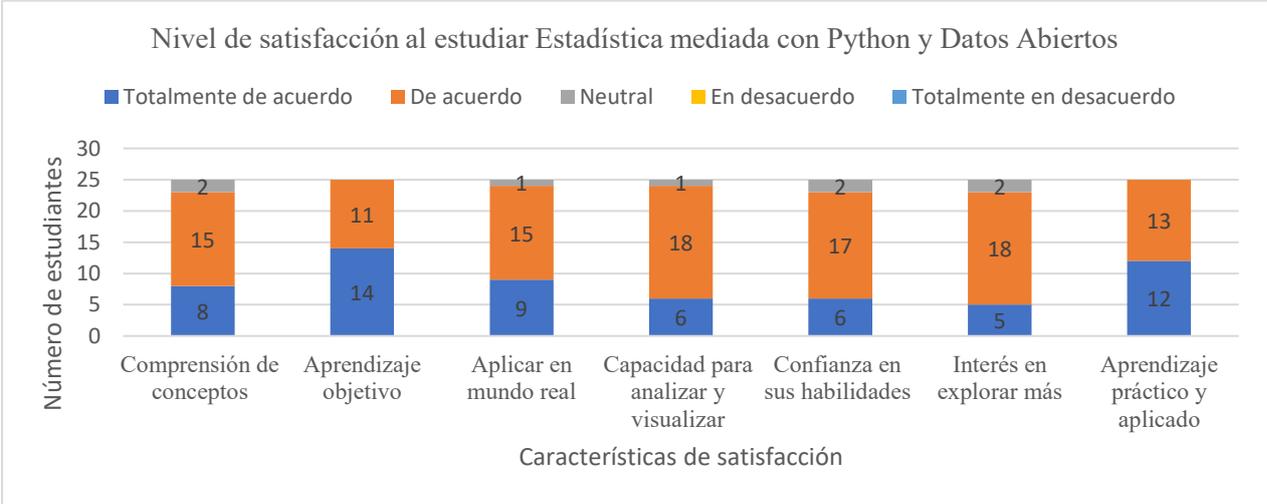
La Tabla 2 presenta aquellos procesos que han sido utilizados por los autores en la enseñanza de la Estadística Descriptiva con datos abiertos y Python. La tabla está distribuida en columnas donde se presentan a los autores, descripción del proceso, el tipo de investigación y por último el resultado de la investigación. Con base en los resultados, se puede evidenciar que el proceso aplicado para enseñar Estadística Descriptiva con datos abiertos y Python es similar en todos los estudios encontrados (Figura 3), generalmente parten con una presentación e introducción a contenidos estadísticos y programación Python, posteriormente se lleva a cabo un taller práctico y por último se trabaja en un proyecto final en el cual los estudiantes aplican lo aprendido en una situación de la vida real. Tales investigaciones han permitido observar los beneficios que brinda la tecnología en la educación y a su vez la contribución a generar un aprendizaje activo del estudiante tanto dentro como fuera del aula.

Resultados de la encuesta

Para dar cumplimiento al segundo objetivo de investigación, que consistió en identificar el grado de satisfacción de los estudiantes del nivel secundaria al estudiar Estadística Descriptiva utilizando datos abiertos gestionados con Python se llevó a cabo una encuesta de 7 preguntas con una escala de satisfacción que va desde, *Totalmente de acuerdo hasta totalmente desacuerdo*. Para la representación de información del segundo objetivo sobre el grado de satisfacción de los estudiantes tras haber aprendido Estadística descriptiva con Python y datos abiertos se construyó la siguiente gráfica de barras múltiple:

Figura 4

Preguntas enfocadas al objetivo específico 2



Nota. Resultados acerca de las preguntas enfocadas en el objetivo específico 2 en la encuesta aplicada.

La gráfica muestra que la mayoría de estudiantes están de acuerdo en que al abordar los tópicos de Estadística como media mediana y moda de la Estadística Descriptiva se mejora la comprensión de estos conceptos, de la misma manera demuestran que esos conceptos se vuelven más objetivos al relacionarlos con datos reales; con respecto a la aplicación en el mundo real, están de acuerdo en que al utilizar bases de datos reales y darles tratamiento con Python han logrado aplicar los conceptos Estadísticos en situaciones reales, a su vez consideran estar de acuerdo en que ha mejorado su capacidad de analizar y visualizar los datos de forma más efectiva.

En relación con la confianza en sus habilidades, los estudiantes están de acuerdo en sentirse más seguros con sus habilidades de Estadística Descriptiva luego de haber utilizado datos reales y trabajados con Python, a la vez la mayoría de ellos están de acuerdo en que ha despertado el interés en profundizar más en los tópicos de esta disciplina, además de ello, consideran estar de acuerdo en que este estudio ha aportado una experiencia de aprendizaje práctico y aplicado a la Estadística.

7. Discusión

La presente investigación analiza el uso de datos abiertos en la enseñanza de la Estadística Descriptiva en el nivel secundario con Python.

El primer cuestionamiento de la investigación se basa en identificar el proceso de enseñanza de la Estadística Descriptiva utilizando datos abiertos con Python, mediante el análisis de diferentes investigaciones se corrobora que los expertos en la enseñanza de esta disciplina siguen una secuencia similar en todos sus estudios, partiendo de una introducción a los conceptos estadísticos, trabajar con ejercicios hipotéticos, presentación de la herramienta tecnológica junto con la familiarización de estas y en ciertas herramientas hace necesario sus bases de programación, posterior a ello en conjunto con los estudiantes llevar a cabo un ejercicio práctico involucrando una base de datos real y como último paso la elaboración de un proyecto final en el cuál van a aplicar los conocimientos y habilidades previamente adquiridas, es decir consolidación de conceptos estadísticos, acercamiento a análisis y conclusiones de bases de datos reales, de tal manera sientan que lo que aprenden es objetivo.

Agregando a lo anterior, en todos los estudios se evidencia que los estudiantes se beneficiaron tras finalizar el estudio, reportaron niveles muy altos de interés y se mostraron especialmente positivos con este tipo de herramientas tecnológicas. Con base en lo antes mencionado, se puede apreciar el valioso instrumento en el que se puede convertir la tecnología en el ámbito educativo dentro de la enseñanza y el aprendizaje de la Estadística, en donde gracias a este tipo de recursos los estudiantes pueden ver a esta asignatura más emocionante e interesante de aprender y no como aquella a la que simplemente deben aprobar por aprobar, tal como lo indican Behar y Ojeada (2016), Alfaro et al. (2015), Quevedo et al. (2015), Rodríguez et al. (2016) y Belfiori (2014).

El segundo cuestionamiento se basa en identificar el grado de satisfacción de los estudiantes del nivel secundaria al estudiar Estadística Descriptiva utilizando datos abiertos gestionados con Python, con base a los resultados se determina satisfactoriamente que los estudiantes consideraron a Python y los datos abiertos como una forma diferente de llevar a cabo el estudio de la Estadística, haciéndola más interesante y práctica. Tal y como lo afirman Blanco (2018), Cuétara et al. (2016), Rodríguez et al. (2016), Medina et al. (2013) y Alpízar (2007) quienes coinciden en que gracias a la tecnología el modo en que se ha llevado a cabo la enseñanza de la Estadística ha cambiado drásticamente, estos autores instan a que un uso adecuado de esta ayuda significativamente al aprendizaje y al modo de pensar y comprender información estadística, además consideran que dentro de la estadística la tecnología permite

desde una mejor comunicación e interacción con la sociedad hasta el desarrollo de nuevos modelos pedagógicos.

Además de ello, Atenas y Havemann (2015), Marín (2022), Medina et al. (2022), Estrella (2017) y Hall (2011) consideran que los datos son como el corazón dentro de la Estadística y más aún cuando son utilizados en la enseñanza de esta disciplina, puesto que los estudiantes al estar expuesto al uso de estos tienen el acceso a la misma información que son utilizadas por varios expertos como científicos responsables políticos, entre otros, lo que les permite enfrentar problemas reales. A su vez, han resaltado la amplia gama de beneficios y desafíos que son exclusivos de la enseñanza y el aprendizaje de Estadística, algunos de ellos son: facilidad de comprensión, mejoran las habilidades estadísticas enormemente y más aún cuando participan en las fases iniciales de planificación de algún proyecto.

La evidencia teórica de las investigaciones muestra que hacer uso de datos abiertos con una herramienta tecnológica, no necesariamente Python si no cualquier otra como Excel, R, SPSS, entre otras, facilita a los estudiantes a adentrarse a la Estadística tal como se encontró en los trabajos de Chang et al. (2016), Schanzer et al. (2022), Tang (2020), Stemock y Kernes (2019) y Vines y Ríos (2022) en donde la motivación por aprender Estadística aumento y por ende obtuvieron como resultado un grado de satisfacción alto, por lo que esto genera argumentos para proponer y diseñar una guía didáctica enmarcada en los trabajos de Duda et al. (2021), Podworny et al. (2022), Lasser et al. (2023) y Fleischer et al. (2022).

8. Conclusiones

Con base en el análisis realizado se concluye que el proceso de enseñanza adecuado para estudiar Estadística Descriptiva con datos abiertos y Python consiste primeramente en una introducción a los conceptos de Estadística, posterior a ello trabajar con ejercicios hipotéticos para luego poder generar la necesidad de hacer uso de las herramientas tecnológicas e introducir los conceptos básicos de programación Python, explorar una base de datos y finalmente la elaboración de un proyecto final, estos pasos se enmarcan en una metodología de Aprendizaje Basado en Proyectos (ABP).

Se concluye que no únicamente con Python se aprende Estadística Descriptiva si no que existen herramientas de interfaz de usuario como de código abierto, en la primera tenemos Excel, JAMOVI, SPSS y en cuanto a introducción de código está R, Python, SAS, entre otras, que también permiten lograr aprendizajes rigurosos en Estadística y motivar a los estudiantes según evidencia científica.

A partir del desarrollo de un conjunto de talleres o clases sobre Estadística Descriptiva orientada con datos abiertos y la herramienta Python los estudiantes logran comprender de mejor manera los conceptos y concluyen satisfactoriamente su proceso con una mayor motivación e interés en profundizar rigurosamente en los tópicos de esta disciplina.

Hay suficiente evidencia científica, documental y empírica para elaborar una propuesta didáctica enmarcada en el Aprendizaje Basado en Proyectos con datos abiertos y el Software estadístico Python.

9. Recomendaciones

Los docentes deben actualizarse constantemente en el uso y manejo de las herramientas y recursos tecnológicos, de tal manera que contribuyan a generar un ambiente innovador durante el proceso de enseñanza y aprendizaje de la Estadística, esto es importante para que puedan aprovechar los beneficios que ofrece los Datos Abiertos y Python.

Para la enseñanza de Estadística Descriptiva es recomendable hacer uso de Softwares Estadísticos y datos reales contextualizados con la mayoría de estudiantes, puesto que permite que ellos vayan consolidando lo aprendido, puedan elaborar su propio conocimiento, concluir sobre situaciones reales, pasar del yo creo al es, de tal modo que se logren aprendizajes significativos.

Seguir la propuesta didáctica que se plantea sobre Aprendizaje Basado en Proyectos (ABP) involucrando Estadística con software para potenciar el aprendizaje de esta disciplina dado que hay evidencia científica de que es una excelente manera de llevar a cabo la enseñanza de la misma, esta propuesta se detalla en el Anexo 1.

10. Bibliografía

- Alayo, A., Soto, E., Hernández, H., y Milagros, M. (2021). Python en la enseñanza de las matemáticas en la carrera de Ingeniería en Ciencias Informáticas. *Serie Científica de la Universidad de las Ciencias Informáticas*, 14(5), 181-202. <https://publicaciones.uci.cu/index.php/serie/article/view/860>
- Alfaro, M., y Alfaro, I. (2018). Uso de la función Solver de Excel para el cálculo de la velocidad de corrosión de acero al carbono en una solución de NaCl al 3,5 % saturada de oxígeno, O₂ Un tutorial práctico. *Educación Química*, 29(2), 17. <https://doi.org/10.22201/fq.18708404e.2018.1.63703>
- Alfaro, N., Campos, S., Ilabaca, F., Ulloa, N. y Moncada, J. (2015). Tic en enseñanza de la estadística descriptiva y aprendizaje autónomo. *Repositorio Digital de documentos en educación de Matemática*, 374-377. <http://funes.uniandes.edu.co/16498/1/Alfaro2015TIC.pdf>
- Alpízar, M. (2007). Herramientas tecnológicas en el proceso de enseñanza y aprendizaje de la estadística. *Cuadernos de Investigación y Formación en Educación Matemática*. <https://revistas.ucr.ac.cr/index.php/cifem/article/view/6893>
- Álvarez, N., Flores, J., y Zambrano, S. (2022). Las Dificultades en la enseñanza - aprendizaje de la estadística y probabilidad: Una perspectiva de estudiantes. *Ecos de la Academia - Universidad Técnica del Norte*, 8(16). <https://doi.org/10.53358/ecosacademia.v8i16.772>
- Améstica, L., King, A., Cornejo, E. y Romero, R. (2019). Aprendizaje activo a través del uso del software Excel en asignaturas de finanzas. *Revista Iberoamericana de Tecnología en Educación y Educación en Tecnología*, 23, 72-79. <https://doi.org/10.24215/18509959.23.e08>
- Arce, M., Conejo, L. y Muñoz, J. (2019). *Aprendizaje y enseñanza de las matemáticas*. https://www.researchgate.net/publication/332471459_Aprendizaje_y_ensenanza_de_las_matematicas
- Atenas, J., y Havemann, L. (2015). *Open Data as Open Educational Resources: Case Studies of Emerging Practice*. <https://doi.org/10.6084/M9.FIGSHARE.1590031>
- Avello, R. y Seisdedo, A. (2017). El procesamiento estadístico con R en la investigación científica. *MediSur*, 15(5), 583-586. <https://medisur.sld.cu/index.php/medisur/article/view/3662>
- Azcárate, P. (2006). ¿Por qué no nos gusta enseñar estadística y probabilidad? In *Conferencia realizada en XII, Jornadas de Investigación en el Aula de Matemáticas: Estadística y azar*,

Granada, España. https://nanopdf.com/download/por-que-no-nos-gusta-ensear-probabilidad-y_pdf#modals

- Ballesteros, E. (2013). La estadística descriptiva como herramienta de análisis en la investigación social. Ejemplo de «actividad práctica» adaptada a la enseñanza del espacio europeo de educación superior. *Números. Revista de Didáctica de las Matemáticas*, 84, 85-98. <https://docta.ucm.es/entities/publication/120109fa-ff65-4d63-914b-e0e155fdb68e>
- Barreto, A. (2012). El progreso de la Estadística y su utilidad en la evaluación del desarrollo. *Papeles de población*, 18(73), 241-271. <https://rppoblacion.uaemex.mx/article/view/8427>
- Batanero, C. (2000). ¿Hacia dónde va la educación estadística? <https://doi.org/10.13140/RG.2.1.3946.7044>
- Batanero, C. (2001). *Didáctica de la Estadística*. <https://doi.org/10.13140/RG.2.1.3946.7044>
- Batanero, C. (2018). *Treinta años de investigación didáctica sobre el análisis inferencial de datos* (pp. 196-209). https://www.researchgate.net/publication/335527780_TREINTA_ANOS_DE_INVESTIGACION_DIDACTICA_SOBRE_EL_ANALISIS_INFERENCIAL_DE_DATOS
- Behar, R., Grima, P., Ojeda, M., y Cruz, C. (2013). *Educación estadística en cursos introductorios a nivel universitario: Algunas reflexiones* (pp. 343-358). https://www.researchgate.net/publication/329058929_Educacion_estadistica_en_cursos_introductorios_a_nivel_universitario_Algunas_reflexiones
- Behar, R., y Ojeda, M. (2016). El papel de algunas opciones tecnológicas en la educación estadística. *Heurística*, 18, 21-35. https://www.researchgate.net/publication/312398845_EL_PAPEL_DE_ALGUNAS OPCIONES TECNOLOGICAS EN LA EDUCACION ESTADISTICA
- Belfiori, L. (2014). *Enseñanza de estadística con recursos TIC*. https://www.researchgate.net/publication/321214003_Ensenanza_de_estadistica_con_recursos_TIC
- Biehler, R., Fleischer, Y., Budde, L., Frischmeier, D., Gerstenberger, D., Podworny, S., y Schulte, C. (2020). Data science education in secondary schools: Teaching and learning decision trees with CODAP and jupyter notebooks as an example of integrating machine learning into statistics education. *New Skills in the Changing World of Statistics Education IASE Roundtable Conference*. New Skills in the Changing World of Statistics Education. <https://doi.org/10.52041/SRAP.20304>

- Blanco, A. (2018). Directrices y recursos para la innovación en la enseñanza de la Estadística en la universidad: Una revisión documental. *REDU. Revista de Docencia Universitaria*, 16(1). <https://doi.org/10.4995/redu.2018.9372>
- Cahyono, D., Naheria, N. y Fauzi, M. S. (2021). Pelatihan Pengolahan Data Penelitian Berbasis Software JASP dan SPSS bagi Mahasiswa FKIP Universitas Mulawarman Kalimantan Timur. *Jurnal Abdi Masyarakat Indonesia*, 1(2). <https://doi.org/10.54082/jamsi.141>
- Calderwood, K. (2002). Incorporating Multiple Epistemologies into Teaching Statistics to Social Work Students. *Journal of Teaching in Social Work*, 22, 17-32. https://doi.org/10.1300/J067v22n01_03
- Cardona, S. (2023). *Efecto del uso de metodología apoyada en las TIC en la enseñanza de la estadística* [Tesis de Maestría, Universidad Nacional de Colombia]. <https://repositorio.unal.edu.co/handle/unal/85455>
- Chang, C., Chin, Y. Y Chang, C. K. (2016). Experimental functionality development for scratch mathematical and statistics extensions. In *2016 International Computer Symposium (ICS)* (pp. 640-644). <https://doi.org/10.1109/ICS.2016.0131>
- Chaves, E. (2016). La enseñanza de la Estadística y la Probabilidad, más allá de procedimientos y técnicas. *Cuadernos de Investigación y Formación en Educación Matemática*, 21-31. <https://revistas.ucr.ac.cr/index.php/cifem/article/view/23880>
- Chaves, S. (2012). *Diseño de una propuesta didáctica para la comprensión de la estadística descriptiva en contextos de demografía* [Tesis de Maestría, Universidad Nacional de Colombia]. <https://repositorio.unal.edu.co/handle/unal/12180>
- Cox, F., González, D., Magreñán, Á., y Orcos, L. (2022). Enseñanza de estadística descriptiva mediante el uso de simuladores y laboratorios virtuales en la etapa universitaria. *Bordón. Revista de Pedagogía*, 74(4), Article 4. <https://doi.org/10.13042/Bordon.2022.94121>
- Cuétara, Y., Salcedo, I. y Hernández, M. (2016). La enseñanza de la estadística: Antecedentes y actualidad en el contexto internacional y nacional. *Atenas*, 3(35), 125-140. <https://atenas.umcc.cu/index.php/atenas/article/view/416>
- Cuevas, J., y Ibáñez, C. (2008). Estándares en educación estadística: Necesidad de conocer la base teórica y empírica que los sustentan. *UNIÓN - REVISTA IBEROAMERICANA DE EDUCACIÓN MATEMÁTICA*, 4(15). <https://union.fespm.es/index.php/UNION/article/view/1141>

- D'Agostino, M., Marti, M., Mejía, F., Cosio, G. y Faba, G. (2018). Estrategia para la gobernanza de datos abiertos de salud: Un cambio de paradigma en los sistemas de información. *Revista Panamericana de Salud Pública*, 41. <https://doi.org/10.26633/rpsp.2017.27>
- Dogan, M. y Can, E. (2019). Jamovi: An Easy to Use Statistical Software for the Social Scientists. *International Journal of Assessment Tools in Education*, 6(4). <https://doi.org/10.21449/ijate.661803>
- Duda, M., Sovacool, K. L., Farzaneh, N., Nguyen, V., Haynes, S., Falk, H., Furman, K., Walker, L., Diao, R., Oneka, M., Drotos, A., Woloshin, A., Dotson, G., Kriebel, A., Meng, L., Thiede, S., Lapp, Z., y Wolford, B. (2021). Teaching Python for Data Science: Collaborative development of a modular & interactive curriculum. *The Journal of open source education*, 4(46), 138. <https://doi.org/10.21105/jose.00138>
- Espinoza, C. y Sánchez, I. (2014). Aprendizaje basado en problemas para enseñar y aprender estadística y probabilidad. *PARADIGMA*, 35(1), 103-128. <http://revistaparadigma.online/ojs/index.php/paradigma/article/view/529>
- Estrella, S. (2017). Enseñar estadística para alfabetizar estadísticamente y desarrollar el razonamiento estadístico. (pp. 173-194). https://www.researchgate.net/publication/316524028_Ensenar_estadistica_para_alfabetizar_estadisticamente_y_desarrollar_el razonamiento estadistico
- Fleischer, Y., Biehler, R. y Schulte, C. (2022). Teaching and learning data-driven machine learning with educationally designed Jupyter Notebooks. *statistics education research journal*, 21(2). <https://doi.org/10.52041/serj.v21i2.61>
- Friel, S. (2007). *The research frontier: Where technology interacts with the teaching and learning of data analysis and statistics*. <https://causeweb.org/cause/research/literature/research-frontier-where-technology-interacts-teaching-and-learning-data-analysis>
- Gea, M., Batanero, C., López, M. y Contreras, J. (2015). Los recursos tecnológicos en la estadística bidimensional en los textos españoles de bachillerato. *Tecné, episteme y didaxis: revista de la Facultad de Ciencia y Tecnología*, 38, 113-132. <https://doi.org/10.17227/01203916.3790>
- Gordon, S. (1995). A Theoretical Approach to Understanding Learners of Statistics. *Journal of Statistics Education*, 3(3). <https://doi.org/10.1080/10691898.1995.11910496>
- Goss, M., y Meneses, J. (2019). *Análisis estadístico con JASP: Una guía para estudiantes*. <https://openaccess.uoc.edu/handle/10609/102926>
- Guajardo, J. (2016). Algunas reflexiones sobre la enseñanza de la estadística. *Revista Académica UC Maule*, 50, 45-48. <https://revistaucmaule.ucm.cl/article/view/17/140>

- Hall, J. (2011). Engaging Teachers and Students with Real Data: Benefits and Challenges. En C. Batanero, G. Burrill, & C. Reading (Eds.), *Teaching Statistics in School Mathematics- Challenges for Teaching and Teacher Education: A Joint ICMI/IASE Study: The 18th ICMI Study* (pp. 335-346). Springer Netherlands. https://doi.org/10.1007/978-94-007-1131-0_32
- Holman, J., y Hacherl, A. (2023). Teaching Monte Carlo Simulation with Python. *Journal of Statistics and Data Science Education*, 31(1), 33-44. <https://doi.org/10.1080/26939169.2022.2111008>
- Lasser, J., Manik, D., Silbersdorff, A., Säfken, B. y Kneib, T. (2021). Introductory data science across disciplines, using Python, case studies, and industry consulting projects. *Teaching Statistics*, 43(1). <https://doi.org/10.1111/test.12243>
- Laureano, D. (2022). Las tecnologías de información y comunicación en la comprensión y producción de textos. *Revista Latinoamericana Ogmios*, 2(3). <https://doi.org/10.53595/rlo.v2.i3.024>
- Lázaro, N., Callejas, Z. y Griol, D. (2022). Utilización del software SPSS para identificar factores predictivos de deserción estudiantil. *Luz*, 21(1), 38-50. <https://luz.uho.edu.cu/index.php/luz/article/view/1155>
- León, N. (2020). Alcances de la enseñanza de la estadística a través de la investigación en la educación media en Venezuela. *PARADIGMA*, 657-684. <https://doi.org/10.37618/PARADIGMA.1011-2251.2020.p657-684.id808>
- LibreTexts (8 de enero de 2024). Unit 4A: Introduction to Statistical Inference. https://stats.libretexts.org/Bookshelves/Applied_Statistics/Biostatistics_-_Open_Learning_Textbook/Unit_4A%3A_Introduction_to_Statistical_Inference
- López, A. (2016). *Usos del programa R en la enseñanza de la estadística*. Comité Latinoamericano de Matemática Educativa. <http://funes.uniandes.edu.co/11865/>
- López, A., Cruañas, J., Salgado, A., Lastayo, L., y Pérez, C. (2015). La enseñanza de la Estadística utilizando herramientas dinámicas computacionales. *Revista Habanera de Ciencias Médicas*, 14(2), 218-226. <https://revhabanera.sld.cu/index.php/rhab/article/view/847>
- Marcano, I. y Benigni, G. (2014). Análisis de alternativas metodológicas para el desarrollo de software educativo. *Saber*, 26(3), 297-304. https://ve.scielo.org/scielo.php?script=sci_arttext&pid=S1315-01622014000300009
- Marín, J. (22 de noviembre de 2022). *Los datos abiertos como herramientas para la educación y la formación*. <https://datos.gob.es/es/blog/los-datos-abiertos-como-herramientas-para-la-educacion-y-la-formacion>

- Marzal, A., Gracia, I., y García, P. (2014). *Introducción a la programación con Python 3*. Universitat Jaume I. <https://doi.org/10.6035/Sapientia93>
- Medina, E., Muñiz, J., Guzmán, D., y Holguín, A. (2022). Recursos y estrategias para la enseñanza de la estadística y la analítica de datos en la educación superior. *Formación universitaria*, 15(3), 61-68. <https://doi.org/10.4067/S0718-50062022000300061>
- Medina, L., Jaquez, J., Noguez, J., García, R. y Tecnológico de Monterrey, C. (2013). *Newton Gymlab: Gimnasio-laboratorio virtual de física y matemáticas*. <https://repositorio.tec.mx/handle/11285/621356>
- Mesa, J. y Caicedo, S. (2020). *Introducción a la estadística descriptiva*. Editorial Universidad de Nariño. <http://sired.udenar.edu.co/6671/>
- Ministerio de Educación. (2016). *Matemática*. MINEDUC https://educacion.gob.ec/wpcontent/uploads/downloads/2016/03/MATE_COMPLETO.pdf
- Moreno, C. y Aguerre, F. (2015). *Estrategia didáctica mediante proyectos formativos para desarrollar capacidades matemáticas en estadística descriptiva en estudiantes del nivel secundario* [Tesis de Maestría, Universidad San Ignacio de Loyola]. <https://hdl.handle.net/20.500.14005/2083>
- Muente, A., y Serale, F. (2018). *Los datos abiertos en América Latina y el Caribe*. <https://doi.org/10.18235/0001202>
- Ortiz, W., Ortega, W., Valencia, L., González, Á. y Gamarra, S. (2021). La educación estadística del ingeniero: Reto de la educación superior. *Revista Universidad y Sociedad*, 13(5), 307-318. http://scielo.sld.cu/scielo.php?script=sci_arttext&pid=S2218-36202021000500307
- Podworny, S., Hüsing, S., y Schulte, C. (2022). A place for a data science project in school: between statistics and epistemic programming. *Statistics education research journal*, 21(2). <https://doi.org/10.52041/serj.v21i2.46>
- Ponteville, C. (2014). *¿Para qué enseñamos estadística?*. Comité Latinoamericano de Matemática Educativa. <http://funes.uniandes.edu.co/5450/>
- Quevedo, D., Gómez, M., y Briseño, M. (2015). Mejora de la enseñanza de la Estadística mediante la implementación de una Comunidad Virtual de Aprendizaje. *Números: Revista de didáctica de las matemáticas*, 90, 73-87. https://www.researchgate.net/publication/363582544_Mejora_de_la_ensenanza_de_la_Estadística_mediante_la_implementacion_de_una_Comunidad_Virtual_de_Aprendizaje

- Ramón, J. y Vilchez, J. (2020). Método clase invertida y desarrollo de competencias estadísticas en estudiantes de maestría. *Revista EDUCARE - UPEL-IPB - Segunda Nueva Etapa 2.0*, 24(3). <https://doi.org/10.46498/reduipb.v24i3.1407>
- Rendón, M., Villasís, M. y Miranda, M. (2016). Estadística descriptiva. *Revista Alergia México*, 63(4), 397-407. <https://revistaalergia.mx/ojs/index.php/ram/article/view/230>
- Rodríguez, J., Romero, J., y Vergara, G. (2016). Importancia de las TIC en enseñanza de las matemáticas. *Revista MATUA* 4(2). <https://investigaciones.uniatlantico.edu.co/revistas/index.php/MATUA/article/view/1861>
- Rodríguez, Y. y Díaz, A. (2009). Herramientas de Minería de Datos. *Revista Cubana de Ciencias Informáticas*, 3(3-4), 73-80. <https://rcci.uci.cu/?journal=rcci&page=article&op=view&path%5B%5D=78>
- Romeu, J. (2013). Uso de software, grupos, proyectos y presentaciones, para enseñar y fomentar la estadística aplicada. *Educación Estadística en América Latina: Tendencias y Perspectivas*, 299-317. <https://web.cortland.edu/romeu/PlantillaAudyEstadInvest2013.pdf>
- Rosabal, D., Martínez, L., y Piquet, N. (2021). Estrategias didácticas para la enseñanza de la Bioestadística, Universidad de Ciencias Médicas de Granma. *cibamanz2021*. <https://cibamanz2021.sld.cu/index.php/cibamanz/cibamanz2021/paper/view/895>
- Ross, S. (2007). *Introducción a la estadística*. Reverte. https://books.google.com.ec/books/about/Introducci%C3%B3n_a_la_estad%C3%ADstica.html?id=pPM2TgQsx8wC&redir_esc=y
- Rouquette, J. y Suárez, A. (2018). La comunicación en la enseñanza de la estadística con el uso de software. *Veredas. Revista del Pensamiento Sociológico*, 35. <https://veredasojs.xoc.uam.mx/index.php/veredas/article/view/467>
- Ruiz, N. (2015). La enseñanza de la Estadística en la Educación Primaria en América Latina. *REICE. Revista Iberoamericana sobre Calidad, Eficacia y Cambio en Educación*, 13(1), 103-121. <https://doi.org/10.15366/reice2015.13.1.006>
- Salinas, J., y Mayén, S. (2016). Estudio exploratorio de las actitudes hacia la estadística en estudiantes mexicanos de bachillerato. *Avances de Investigación en Educación Matemática*, 10. <https://doi.org/10.35763/aiem.v0i10.130>
- Santabárbara, J. y Lasheras, I. (2020). Docencia de Bioestadística en Medicina con software gratuito jamovi: Una ventana de oportunidad. *Revista Española de Educación Médica*, 1(1). <https://doi.org/10.6018/edumed.421421>

- Santoyo, F. (2022). La enseñanza de la estadística en el contexto de la sociedad del dato: desafíos y reflexiones. *Journal De Ciencias Sociales*, 1(18), 88-106. <https://doi.org/10.18682/jcs.vi18.4338>
- Schanzer, E., Pfenning, N., Denny, F., Dooman, S., Politz, J., Lerner, B., Fisler, K., y Krishnamurthi, S. (2022). Integrated Data Science for Secondary Schools: Design and Assessment of a Curriculum. *Proceedings of the 53rd ACM Technical Symposium on Computer Science Education*, 22-28. <https://doi.org/10.1145/3478431.3499311>
- StataCorp. (2023). *Stata Statistical Software: Release 18*. College Station, TX: StataCorp LLC. <https://www.stata.com/>
- Stemock, B., y Kerns, L. (2019). Use of commercial and free software for teaching statistics. *STATISTICS EDUCATION RESEARCH JOURNAL*, 18(2). <https://doi.org/10.52041/serj.v18i2.140>
- Suasnabas, L. y Juárez, J. (2020). Calidad de la educación en Ecuador. ¿Mito o realidad? *Dominio de las Ciencias*, 6(2). <https://doi.org/10.23857/dc.v6i2.1160>
- Tang, M. (2020). A simple introduction to programming and statistics with decision trees in R. *Teaching Statistics*, 42(2), 36-40. <https://doi.org/10.1111/test.12210>
- Thorsby, J., Stowers, G., Wolslegel, K., y Tumbuan, E. (2017). Understanding the content and features of open data portals in American cities. *Government Information Quarterly*, 34(1), 53-61. <https://doi.org/10.1016/j.giq.2016.07.001>
- Trujillo, R., Hernández, A., Bueno, A. y Palacio, R. (2015). Consideraciones sobre el uso de los paquetes estadísticos en la enseñanza de la asignatura de Estadística en la carrera de Medicina. *Revista Información Científica*, 92(4), 864-873. <https://revinformcientifica.sld.cu/index.php/ric/article/view/205/1377>
- Vila, K. (2018). *Datos abiertos para el desarrollo de competencias en Educación Secundaria y Formación Profesional* [Tesis de maestría, Universidad de Alicante]. <http://rua.ua.es/dspace/handle/10045/76547#vpreview>
- Vinces, F. y Ríos, W. (2022). Cuadernos interactivos: experiencia de enseñanza y aprendizaje de estadística en el entorno Jupyter. *Avances en Matemática Educativa*
- Wild, C., Utts, J., y Horton, N. (2018). What Is Statistics? En D. Ben-Zvi, K. Makar, y J. Garfield (Eds.), *International Handbook of Research in Statistics Education* (pp. 5-36). Springer International Publishing. https://doi.org/10.1007/978-3-319-66195-7_1

- Wilson, B., y Cong, C. (2021). Beyond the supply side: Use and impact of municipal open data in the U.S. *Telematics and Informatics*, 58, 101526. <https://doi.org/10.1016/j.tele.2020.101526>
- Wistuba, L. (2014). Didáctica de la estadística: modelos culturales en la enseñanza de la estadística. *Sabvia Revista de Educación*, 1, 1-30. <https://funes.uniandes.edu.co/funes-documentos/didactica-de-la-estadistica-modelos-culturales-en-la-ensenanza-de-la-estadistica/>
- Zuiderwijk, A., Janssen, M. y Susha, I. (2015). Improving the Speed and Ease of Open Data Use Through Metadata, Interaction Mechanisms and Quality Indicators. *Journal of Organizational Computing and Electronic Commerce*, 26. <https://doi.org/10.1080/10919392.2015.1125180>
- Zuluaga, J., y Villa, J. (2018). *Percepciones de estudiantes acerca de una estrategia basada en la Web 2.0 en un curso de estadística general*. <https://www.semanticscholar.org/paper/Percepciones-de-estudiantes-acerca-de-una-basada-en-Zuluaga-Arango-Villa-Ochoa/6ed973bad69436e8e548a2e1f793af36a35b0665>
- Zúñiga, F. y Diaz, D. (2017). *Implementación de una secuencia didáctica basada en situaciones problema para comprender la comprensión de la noción de aleatoriedad* [Tesis de Maestría, Universidad del Valle]. <https://hdl.handle.net/10893/10528>

Anexo 1. Propuesta Didáctica

PROPUESTA DIDÁCTICA

*Propuesta didáctica innovadora: Enseñanza de la Estadística
Descriptiva con datos abiertos mediada con Python*

ÍNDICE

| | |
|--|-----|
| TÍTULO DE LA PROPUESTA..... | 54 |
| PRESENTACIÓN..... | 55 |
| OBJETIVOS | 56 |
| JUSTIFICACIÓN | 57 |
| DESARROLLO | 64 |
| Clase 1: Introducción a la Estadística y tipos de variables | 64 |
| Clase 2: Tratamiento de variables y técnicas de visualización de datos | 69 |
| Clase 3: Medidas de dispersión y visualización de Datos..... | 79 |
| Clase 4: Introducción a datos abiertos y Python | 85 |
| Clase 5: Proyecto guiado de análisis de datos con Python (Anexo 1.) | 88 |
| Clase 6: Proyecto conjunto de análisis de datos con Python (Anexo 2.) | 94 |
| RESULTADOS ESPERADOS | 101 |
| BIBLIOGRAFIA..... | 102 |
| ANEXOS..... | 103 |

TÍTULO DE LA PROPUESTA

Propuesta didáctica innovadora: Enseñanza de la Estadística Descriptiva con datos abiertos mediada con Python

La Estadística es una disciplina clave para interpretar y analizar datos de manera efectiva. Esta propuesta didáctica se centra en la enseñanza de conceptos fundamentales de la Estadística Descriptiva utilizando datos abiertos y herramienta de programación Python. Se busca que los estudiantes de nivel secundario no solo comprendan los conceptos teóricos, sino que también desarrollen habilidades prácticas aplicables en el análisis de datos reales. El proceso de enseñanza está diseñado para ser implementado en un curso de Estadística Descriptiva para estudiantes de 2do de Bachillerato General Unificado (BGU) y consta de las siguientes etapas basadas en la metodología de Vines y Ríos (2022): Inicio (Introducción a Conceptos Básicos de Estadística, Ejercicios Hipotéticos, Generar Necesidad de Utilizar Herramienta Tecnológica), Desarrollo (Introducción a Programación Python, Exploración de Datos a Trabajar) y Cierre (Proyecto Final)

Esta estructura permite un enfoque progresivo, donde los estudiantes primero adquieren una base teórica sólida, luego practican con datos ficticios, y finalmente aplican sus conocimientos en un proyecto práctico con datos reales. Al finalizar el curso, los estudiantes habrán desarrollado una comprensión profunda de la Estadística Descriptiva y habilidades prácticas en programación y análisis de datos, preparándolos para futuros desafíos académicos y profesionales.

La propuesta está estructurada de la siguiente manera: portada, índice, título, presentación; objetivos que guían el desarrollo de la propuesta; justificación que explica las razones para adoptar esta metodología; planificación detallada de las etapas del proceso educativo; desarrollo donde se describe el procedimiento de cada clase; esquema del proyecto; resultados esperados, que delinear las metas del proyecto y finalmente, la bibliografía.

OBJETIVOS

Objetivo General:

- Fortalecer la enseñanza de la Estadística Descriptiva en el nivel secundario utilizando datos abiertos y herramientas de programación en Python, con el fin de mejorar la comprensión y aplicación de los conceptos estadísticos.

Objetivos Específicos:

- Elaborar una Planificación micro curricular para la enseñanza y aprendizaje de la Estadística Descriptiva incorporando herramientas como Python y datos abiertos.
- Desarrollar las actividades de aprendizaje para demostrar el uso de datos abiertos gestionados con Python.

JUSTIFICACIÓN

La propuesta de integrar datos abiertos y herramientas de programación como Python en la enseñanza de la Estadística Descriptiva en el nivel secundario se fundamenta en la necesidad de modernizar los métodos educativos y preparar a los estudiantes para los desafíos del siglo XXI. En un entorno cada vez más digitalizado, es crucial que los estudiantes adquieran habilidades prácticas en el manejo y análisis de datos, habilidades que son fundamentales en numerosas disciplinas académicas y profesionales. Python, con sus potentes bibliotecas como Pandas, Seaborn, Matplotlib, entre otras, que ofrecen una plataforma accesible y robusta para explorar conceptos estadísticos de manera efectiva, permitiendo a los estudiantes realizar análisis complejos y visualizaciones de datos de manera interactiva y comprensible.

Además de promover competencias digitales esenciales, esta propuesta también fomenta un aprendizaje activo y colaborativo a través del Aprendizaje Basado en Proyectos (ABP), dado que este método evidencia mayor efectividad y uso. Al trabajar con datos reales y enfrentarse a problemas prácticos, los estudiantes no solo consolidan su comprensión teórica de la estadística descriptiva, sino que también desarrollan habilidades cruciales como el trabajo en equipo, la resolución de problemas y la comunicación efectiva de resultados. El acceso a datos abiertos enriquece su experiencia educativa al permitirles explorar temas relevantes y actuales en diversos campos, preparándolos para aplicar sus conocimientos en contextos del mundo real y para futuros estudios y carreras donde el análisis de datos es fundamental

PLANIFICACIÓN MICROCURRICULAR

DATOS INFORMATIVOS

| | | | |
|--------------------|--------------------------------------|---------------------|---------|
| DOCENTE(S): | Jean Prado | GRADO/CURSO: | Segundo |
| ÁREA: | Matemática | PARALELO(S): | “A” |
| ASIGNATURA: | Matemáticas | TRIMESTRE: | Tercero |
| SUBNIVEL: | Bachillerato General Unificado (BGU) | | |

APRENDIZAJE DISCIPLINAR E INTERDISCIPLINAR

| OBJETIVO DE APRENDIZAJE: | <p>O.M.5.4. Valorar el empleo de las TIC para realizar cálculos y resolver, de manera razonada y crítica, problemas de la realidad nacional, argumentando la pertinencia de los métodos utilizados y juzgando la validez de los resultados.</p> <p>O.M.5.6. Desarrollar la curiosidad y la creatividad a través del uso de herramientas matemáticas al momento de enfrentar y solucionar problemas de la realidad nacional, demostrando actitudes de orden, perseverancia y capacidades de investigación.</p> | | | | |
|---|---|---|---|---|---|
| DESTREZAS CON CRITERIO DE DESEMPEÑO | INDICADORES DE EVALUACIÓN | ESTRATEGIAS METODOLÓGICAS ACTIVAS PARA LA ENSEÑANZA APRENDIZAJE | RECURSOS | ACTIVIDADES EVALUATIVAS | |
| | | | | TÉCNICAS | INSTRUMENTOS |
| <p>M.5.3.1. Calcular e interpretar la media, mediana, moda, rango, varianza y desviación estándar para datos no agrupados y agrupados, con apoyo de las TIC.</p> <p>M.5.3.2. Resolver y plantear problemas de aplicación de las medidas de tendencia central y de dispersión para datos agrupados, con apoyo de las TIC.</p> <p>M.5.3.3. Juzgar la validez de las soluciones obtenidas en los problemas de aplicación de las medidas de tendencia central y de</p> | <p>I.M.5.9.1. Calcular, con y sin apoyo de las TIC, las medidas de centralización y dispersión para datos agrupados y no agrupados; representa la información en gráficos estadísticos apropiados y los interpreta, juzgando su validez. (J.2.,I.3.)</p> | <p style="text-align: center;">Fase Inicial</p> <p>Clase 1: Introducción a la Estadística y Tipos de Variables</p> <ul style="list-style-type: none"> • Saludo de bienvenida. • Tomar asistencia. • Presentación de la agenda del día. • Introducción a la Estadística • Definición de la Estadística • Importancia de la Estadística • Introducción al concepto de variable. • Tipos de Variables • Variables cualitativas: Ordinal y nominal • Variables cuantitativas: Discreta y continua | <ul style="list-style-type: none"> • Pizarra. • Marcadores borrador y • Reglas • Diapositivas | <ul style="list-style-type: none"> • Observación. • Cuestionarios. • Rúbrica. • Talleres. | <ul style="list-style-type: none"> • Guía de observación. • Registro anecdótico. • Diario de clases. • Trabajos grupales. • Test o evaluaciones. |

| | | | | | |
|---|--|---|---|---|---|
| <p>dispersión para datos agrupados dentro del contexto del problema, con apoyo de las TIC.</p> | | | | | |
| <p>M.5.3.1. Calcular e interpretar la media, mediana, moda, rango, varianza y desviación estándar para datos no agrupados y agrupados, con apoyo de las TIC.</p> <p>M.5.3.2. Resolver y plantear problemas de aplicación de las medidas de tendencia central y de dispersión para datos agrupados, con apoyo de las TIC.</p> <p>M.5.3.3. Juzgar la validez de las soluciones obtenidas en los problemas de aplicación de las medidas de tendencia central y de dispersión para datos agrupados dentro del contexto del problema, con apoyo de las TIC.</p> | <p>I.M.5.9.1. Calcula, con y sin apoyo de las TIC, las medidas de centralización y dispersión para datos agrupados y no agrupados; representa la información en gráficos estadísticos apropiados y los interpreta, juzgando su validez. (J.2.,I.3.)</p> | <p>Clase 2: Tratamiento de Variables y Técnicas de Visualización</p> <ul style="list-style-type: none"> • Saludo de bienvenida. • Tomar asistencia. • Presentación de la agenda del día. • Introducción al tratamiento de variables tanto cualitativas como cuantitativas • Variables cualitativas: distribución de frecuencias. • Variable cuantitativa: Medidas de tendencia central y técnicas de visualización de datos. | <ul style="list-style-type: none"> • Pizarra. • Marcadores y borrador • Reglas • Diapositivas | <ul style="list-style-type: none"> • Observación. • Cuestionarios. • Rúbrica. • Talleres. | <ul style="list-style-type: none"> • Guía de observación. • Registro anecdótico. • Diario de clases. • Trabajos grupales. • Test o evaluaciones. |
| <p>M.5.3.1. Calcular e interpretar la media, mediana, moda, rango, varianza y desviación estándar para datos no agrupados y agrupados, con apoyo de las TIC.</p> <p>M.5.3.2. Resolver y plantear problemas de aplicación de las medidas</p> | <p>I.M.5.9.1. Calcula, con y sin apoyo de las TIC, las medidas de centralización y dispersión para datos agrupados y no agrupados; representa la información en gráficos estadísticos apropiados y los interpreta,</p> | <p>Clase 3: Medidas de Dispersión y Visualización de Datos</p> <ul style="list-style-type: none"> • Saludo de bienvenida. • Tomar asistencia. • Presentación de la agenda del día. • Medidas de dispersión de datos | <ul style="list-style-type: none"> • Pizarra. • Marcadores y borrador • Reglas | <ul style="list-style-type: none"> • Observación. • Cuestionarios. • Rúbrica. • Talleres. | <ul style="list-style-type: none"> • Guía de observación. • Registro anecdótico. • Diario de clases. • Trabajos grupales. |

| | | | | | |
|---|--|---|--|---|---|
| <p>de tendencia central y de dispersión para datos agrupados, con apoyo de las TIC.</p> <p>M.5.3.3. Juzgar la validez de las soluciones obtenidas en los problemas de aplicación de las medidas de tendencia central y de dispersión para datos agrupados dentro del contexto del problema, con apoyo de las TIC.</p> | <p>juzgando su validez. (J.2.,I.3.)</p> | <ul style="list-style-type: none"> • Técnicas de visualización de datos • Cierre de la clase • Mención sobre datos reales | <ul style="list-style-type: none"> • Diapositivas | | <ul style="list-style-type: none"> • Test o evaluaciones. |
| <p>M.5.3.1. Calcular e interpretar la media, mediana, moda, rango, varianza y desviación estándar para datos no agrupados y agrupados, con apoyo de las TIC.</p> <p>M.5.3.2. Resolver y plantear problemas de aplicación de las medidas de tendencia central y de dispersión para datos agrupados, con apoyo de las TIC.</p> <p>M.5.3.3. Juzgar la validez de las soluciones obtenidas en los problemas de aplicación de las medidas de tendencia central y de dispersión para datos agrupados dentro del contexto del problema, con apoyo de las TIC.</p> | <p>I.M.5.9.1. Calcula, con y sin apoyo de las TIC, las medidas de centralización y dispersión para datos agrupados y no agrupados; representa la información en gráficos estadísticos apropiados y los interpreta, juzgando su validez. (J.2.,I.3.)</p> | <p style="text-align: center;">Fase de desarrollo</p> <p>Clase 4: Introducción a Datos Abiertos y Python</p> <ul style="list-style-type: none"> • Saludo de bienvenida. • Tomar asistencia. • Presentación de la agenda del día. • Datos abiertos: Definición e importancia. • Introducción a Python • ¿Qué es Python? • Importancia de Python | <ul style="list-style-type: none"> • Pizarra. • Software Python • Bases de Datos • Marcadores y borrador • Reglas • Diapositivas | <ul style="list-style-type: none"> • Observación. • Cuestionarios. • Rúbrica. • Talleres. | <ul style="list-style-type: none"> • Guía de observación. • Registro anecdótico. • Diario de clases. • Trabajos grupales. • Test o evaluaciones. |

| | | | | | |
|---|--|--|--|---|---|
| <p>M.5.3.1. Calcular e interpretar la media, mediana, moda, rango, varianza y desviación estándar para datos no agrupados y agrupados, con apoyo de las TIC.</p> <p>M.5.3.2. Resolver y plantear problemas de aplicación de las medidas de tendencia central y de dispersión para datos agrupados, con apoyo de las TIC.</p> <p>M.5.3.3. Juzgar la validez de las soluciones obtenidas en los problemas de aplicación de las medidas de tendencia central y de dispersión para datos agrupados dentro del contexto del problema, con apoyo de las TIC.</p> | <p>I.M.5.9.1. Calcula, con y sin apoyo de las TIC, las medidas de centralización y dispersión para datos agrupados y no agrupados; representa la información en gráficos estadísticos apropiados y los interpreta, juzgando su validez. (J.2.,I.3.)</p> | <p>Clase 5: Proyecto Guiado de Análisis de Datos con Python</p> <ul style="list-style-type: none"> • Saludo de bienvenida. • Tomar asistencia. • Presentación de la agenda del día. • Introducción al proyecto guiado: Explicación de la importancia del proyecto y de cada etapa del análisis de datos. • Importación de Librerías: Explicación y ejecución del código para importar librerías necesarias (Pandas, Matplotlib). • Importación de Datos: Cargar datos desde un archivo CSV y mostrar cómo visualizar los primeros registros. • Limpieza y Preparación de Datos: Mostrar cómo manejar valores nulos y preparar los datos para el análisis. • Medidas de Tendencia Central: Calcular moda, mediana y media. • Medidas de Dispersión: Calcular varianza y desviación estándar. • Visualización de Datos: Crear histogramas y gráficos de barras. • Cierre de la clase: Resumen de lo aprendido; Preguntas y respuestas. | <ul style="list-style-type: none"> • Pizarra. • Software Python • Bases de Datos • Marcadores y borrador • Reglas • Diapositivas | <ul style="list-style-type: none"> • Observación. • Cuestionarios. • Rúbrica. • Talleres. | <ul style="list-style-type: none"> • Guía de observación. • Registro anecdótico. • Diario de clases. • Trabajos grupales. • Test o evaluaciones. |
| <p>M.5.3.1. Calcular e interpretar la media,</p> | <p>I.M.5.9.1. Calcula, con y sin apoyo de las TIC,</p> | <p>Fase Final</p> | | <ul style="list-style-type: none"> • Observación. | <ul style="list-style-type: none"> • Guía de observación. |

| | | | | | |
|--|---|---|--|---|---|
| <p>mediana, moda, rango, varianza y desviación estándar para datos no agrupados y agrupados, con apoyo de las TIC.</p> <p>M.5.3.2. Resolver y plantear problemas de aplicación de las medidas de tendencia central y de dispersión para datos agrupados, con apoyo de las TIC.</p> <p>M.5.3.3. Juzgar la validez de las soluciones obtenidas en los problemas de aplicación de las medidas de tendencia central y de dispersión para datos agrupados dentro del contexto del problema, con apoyo de las TIC.</p> | <p>las medidas de centralización y dispersión para datos agrupados y no agrupados; representa la información en gráficos estadísticos apropiados y los interpreta, juzgando su validez. (J.2.,I.3.)</p> | <p>Clase 6: Proyecto Conjunto de Análisis de Datos con Python</p> <ul style="list-style-type: none"> • Saludo de bienvenida. • Tomar asistencia. • Presentación de la agenda del día. • Introducción al proyecto conjunto: Explicación de la nueva base de datos y su relevancia. • Importación de Librerías y Datos: Juntos, importar librerías y cargar la nueva base de datos. • Limpieza y Preparación de Datos: preparar los datos con la participación activa de los estudiantes. • Medidas de Tendencia Central: Calcular moda, mediana y media con la guía del docente y la colaboración de los estudiantes. • Medidas de Dispersión: Calcular varianza y desviación estándar de forma conjunta. • Visualización de Datos: Crear histogramas y gráficos de barras con la participación de los estudiantes. • Cierre de la clase y del curso: Presentación y discusión de resultados; Evaluación del proyecto y reflexión sobre lo aprendido; Preguntas y respuestas finales. | <ul style="list-style-type: none"> • Pizarra. • Software Python • Bases de Datos • Marcadores y borrador • Reglas • Diapositivas | <ul style="list-style-type: none"> • Cuestionarios. • Rúbrica. • Talleres. | <ul style="list-style-type: none"> • Registro anecdótico. • Diario de clases. • Trabajos grupales. • Test o evaluaciones. |
|--|---|---|--|---|---|

PROYECTO INTEGRADOR

NOMBRE DEL PROYECTO INTERDISCIPLINAR:

| FASE: | | | | | |
|---|---------------------------|---|----------|-------------------------|--------------|
| ÁREAS ACADÉMICAS QUE SE VINCULAN: | | | | | |
| OBJETIVO DE APRENDIZAJE: | | | | | |
| PRODUCTO DE LA FASE: | | | | | |
| DESTREZAS CON CRITERIOS DE DESEMPEÑO | INDICADORES DE EVALUACIÓN | ESTRATEGIAS METODOLÓGICAS ACTIVAS PARA LA ENSEÑANZA APRENDIZAJE | RECURSOS | ACTIVIDADES EVALUATIVAS | |
| | | | | TÉCNICAS | INSTRUMENTOS |
| | | | | | |
| ADAPTACIONES CURRICULARES PARA ESTUDIANTES CON NECESIDADES ESPECÍFICAS | | | | | |
| DESTREZAS CON CRITERIO DE DESEMPEÑO | INDICADORES DE EVALUACIÓN | ESTRATEGIAS METODOLÓGICAS ACTIVAS PARA LA ENSEÑANZA APRENDIZAJE | RECURSOS | EVALUACIÓN | |
| | | | | TÉCNICAS | INSTRUMENTOS |
| | | | | | |
| ELABORADO POR: | | REVISADO POR: | | APROBADO POR: | |
| _____ | | | | | |
| Jean Prado | | | | | |
| Fecha: | Fecha: | Fecha: | | Fecha: | |
| | | | | | |

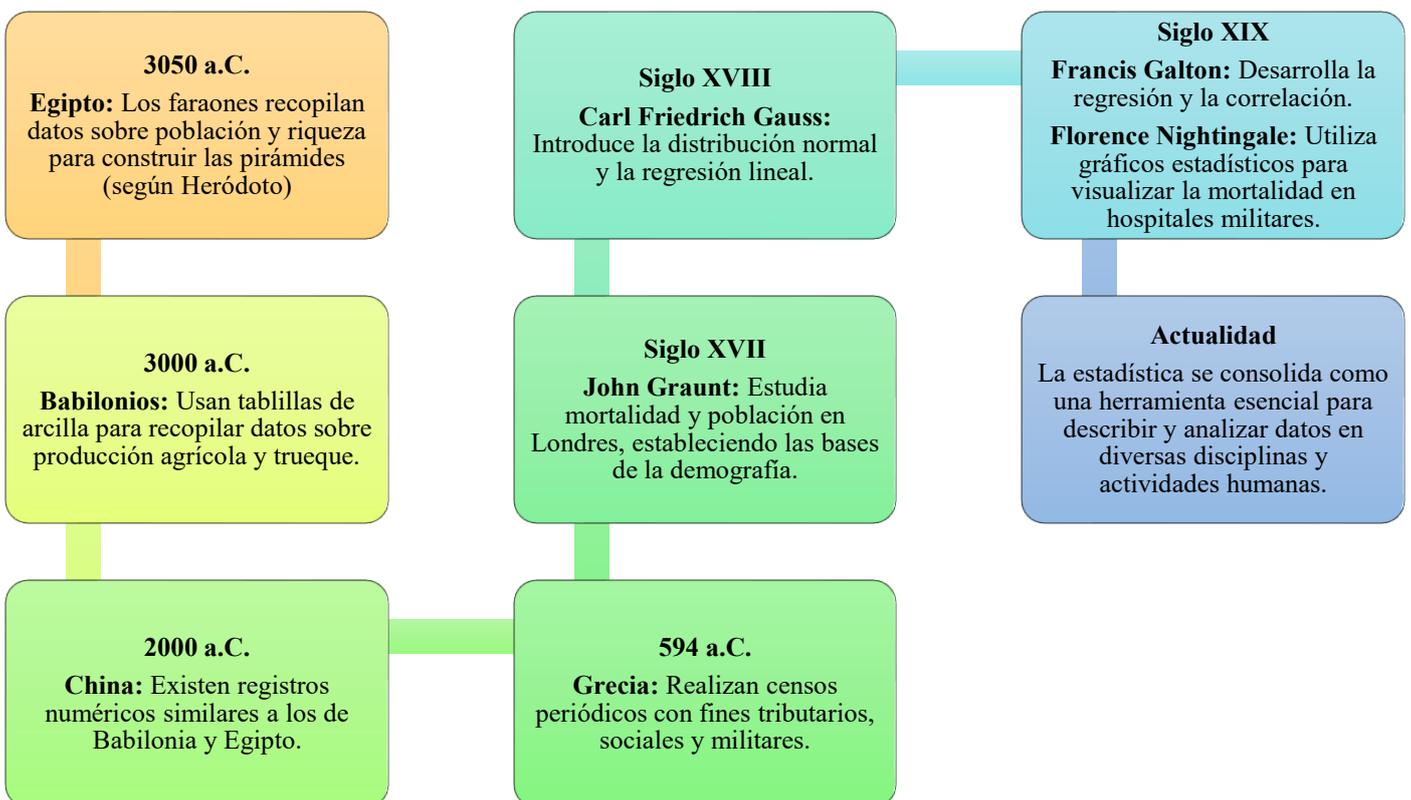
DESARROLLO

Nota: Las imágenes encontradas en la propuesta fueron obtenidas de *Pixibay*

| Clase 1 | |
|--------------------------------|---|
| Asignatura | Matemática |
| Curso | 2do de BGU |
| Tema | Introducción a la Estadística y tipos de variables |
| Destreza | <p>M.5.3.1. Calcular e interpretar la media, mediana, moda, rango, varianza y desviación estándar para datos no agrupados y agrupados, con apoyo de las TIC.</p> <p>M.5.3.2. Resolver y plantear problemas de aplicación de las medidas de tendencia central y de dispersión para datos agrupados, con apoyo de las TIC.</p> <p>M.5.3.3. Juzgar la validez de las soluciones obtenidas en los problemas de aplicación de las medidas de tendencia central y de dispersión para datos agrupados dentro del contexto del problema, con apoyo de las TIC.</p> |
| Indicador de evaluación | I.M.5.9.1. Calcula, con y sin apoyo de las TIC, las medidas de centralización y dispersión para datos agrupados y no agrupados; representa la información en gráficos estadísticos apropiados y los interpreta, juzgando su validez. (J.2.,I.3.) |

PARTE TEÓRICA

Historia de la Estadística



Estadística: Es la ciencia que se encarga de la recolección, ordenamiento, representación, análisis e interpretación de datos generados en una investigación sobre hechos, individuos o grupos de los mismos, para deducir de ello conclusiones precisas o estimaciones futuras.



Importancia de la Estadística:

Los métodos estadísticos se utilizan para propósitos descriptivos, para organizar y resumir datos numéricos. Ahora bien, las técnicas estadísticas se aplican de manera amplia en mercadotecnia, contabilidad, control de calidad, estudios de consumidores, análisis de resultados en deportes, administración de instituciones, en la educación, organismos políticos, en la medicina y en otras muy distintas áreas como un auxiliar en la toma de decisiones.

Introducción al concepto de variable:

Ahora que hemos visto la definición e importancia de la estadística, es esencial entender que la estadística se basa en el análisis de ‘variables’.

Una variable es una característica que cambia o varía con el tiempo y/o para diferentes personas u objetos bajo consideración.



En estadística, estas variables nos permiten observar patrones, medir y cuantificar los datos, y finalmente, hacer inferencias o predicciones basadas en los datos. Ejemplos:

Edad de los estudiantes en una clase:

Imagina que estamos analizando datos de una escuela. La variable “edad” representa cuántos años tienen los estudiantes. Algunos pueden tener 15 años, otros 18 o 20. Observar esta variable nos permite identificar patrones de distribución de edades en la población estudiantil.





Altura de los jugadores de un equipo de baloncesto:

Supongamos que estamos investigando un equipo de baloncesto profesional. La variable “altura” describe cuán altos son los jugadores. Algunos pueden medir 6 pies y 2 pulgadas (188 cm), mientras que otros pueden ser más altos o más bajos. Esta variable es importante para evaluar las características físicas del equipo.

Peso de los pacientes en un hospital:

Consideremos que estamos analizando datos médicos. La variable “peso” representa cuántos kilogramos o libras pesan los pacientes. Algunos pueden pesar 70 kg, otros 80 kg o más. Esta información es relevante para el tratamiento médico y la gestión hospitalaria.



A continuación, vamos a explorar los diferentes tipos de variables que encontramos en estadística y cómo cada tipo se utiliza en el análisis de datos.

Tipos de variables

De tal manera los datos que se generan en un estudio estadístico, serán de la misma categorización de la variable que se está estudiando, por lo tanto, éstos pueden ser cualitativos o cuantitativos.

Variable cualitativa

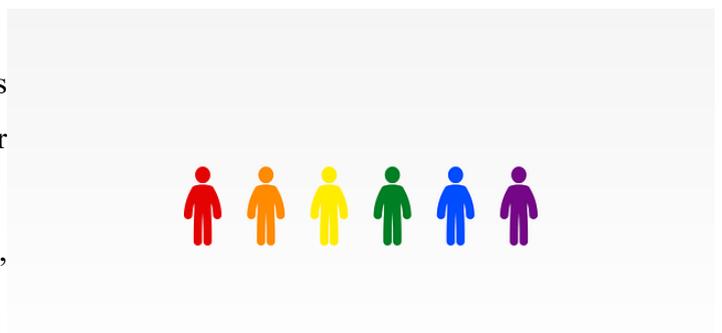
Las variables cualitativas se refieren a características o cualidades que no pueden ser medidas con números.

Podemos distinguir dos tipos:

Variable cualitativa nominal

Una variable cualitativa nominal presenta modalidades no numéricas que no admiten un criterio de orden. Por ejemplo:

- **Género:** Las categorías son “masculino”, “femenino” y “otro”.



- **Nacionalidad:** Las categorías pueden ser “español”, “mexicano”, “colombiano”, etc.
- **Tipo de comida preferida:** Las categorías pueden ser “china”, “mexicana”, “italiana”, etc.



Variable cualitativa ordinal

Una variable cualitativa ordinal presenta modalidades no numéricas, en las que existe un orden. Por ejemplo:

- **La nota en un examen:** suspenso, aprobado, notable, sobresaliente.
- **Puesto conseguido en una prueba deportiva:** 1º, 2º, 3º, ...
- **Medallas de una prueba deportiva:** oro, plata, bronce

Variable cuantitativa

Una variable cuantitativa es la que se expresa mediante un número, por tanto, se pueden realizar operaciones aritméticas con ella.

Podemos distinguir dos tipos:

Variable discreta

Una variable discreta es aquella que toma valores aislados, es decir no admite valores intermedios entre dos valores específicos. Por ejemplo:

- **El número de hermanos de 5 amigos:** 2, 1, 0, 1, 3.
- **Cantidad de estudiantes en una clase:** 20, 35, 25, 40
- **Número de automóviles en un estacionamiento en un momento dado:** 2, 4, 5, 7



Variable continúa

Una variable continua es aquella que puede tomar valores comprendidos entre dos números. Por ejemplo:

- **La altura de los 5 amigos:** 1.73, 1.82, 1.77, 1.69, 1.75.
- **Distancia recorrida por un automóvil deportivo en una pista:** 2.3 km, 1.5 km, 5.2 km, 1.3 km
- **Volumen de una pieza metálica:** 4.5, 3.2, 5.9, 6.4, 2.3



¿Sabías que...?

El Ecuador realiza censos nacionales aproximadamente cada 10 años. El censo más reciente se llevó a cabo en 2022. Estos censos son esenciales para obtener información detallada sobre la población, como la distribución por edades, género, nivel educativo, y más, permitiendo a los gobiernos planificar y asignar recursos de manera más efectiva.



Curiosidades Estadísticas en Ecuador:

El turismo es una industria importante en Ecuador. Según estadísticas recientes, el país recibe millones de turistas cada año, atraídos por sus bellezas naturales como las Islas Galápagos, la Amazonía, y los Andes. Las estadísticas de turismo ayudan a planificar y mejorar las infraestructuras y servicios turísticos.

| Clase 2 | |
|--|---|
| Asignatura | Matemática |
| Curso | 2do de BGU |
| Tema | Tratamiento de variables y técnicas de visualización de datos |
| Destreza  | <p>M.5.3.1. Calcular e interpretar la media, mediana, moda, rango, varianza y desviación estándar para datos no agrupados y agrupados, con apoyo de las TIC.</p> <p>M.5.3.2. Resolver y plantear problemas de aplicación de las medidas de tendencia central y de dispersión para datos agrupados, con apoyo de las TIC.</p> <p>M.5.3.3. Juzgar la validez de las soluciones obtenidas en los problemas de aplicación de las medidas de tendencia central y de dispersión para datos agrupados dentro del contexto del problema, con apoyo de las TIC.</p> |
| Indicador de evaluación | I.M.5.9.1. Calcula, con y sin apoyo de las TIC, las medidas de centralización y dispersión para datos agrupados y no agrupados; representa la información en gráficos estadísticos apropiados y los interpreta, juzgando su validez. (J.2.,I.3.) |

PARTE TEÓRICA

Tratamiento de variables Cualitativas.

Distribución de frecuencias: Es una tabla estadística donde se presentan los datos resumidos, de tal manera que se puede en una visión panorámica establecer un criterio sobre su comportamiento, entendiéndose por comportamiento, la determinación aproximada de los valores centrales, la variabilidad que presentan y si son o no relativamente simétricos con relación a un valor central.

En una tabla de frecuencias se pueden resumir cualquier tipo de datos, categóricos (nominales), ordinales, discretos. Para los tipos de datos, es decir, nominales, ordinales y discretos, la distribución de frecuencias constará básicamente de dos columnas, la izquierda reservada para las categorías (valores) que aparecen en el estudio respectivo y la de la derecha donde se ubica la frecuencia de clase respectiva



Ejemplo de Distribución de Frecuencias para una Variable Nominal: Preferencias de Género

Musical

Supongamos que hemos encuestado a 50 personas en una universidad para conocer sus preferencias de género musical. Las opciones posibles son: Rock, Pop, Jazz, Electrónica y Clásica.



- **Datos:**

Se han recopilado las preferencias de género musical de 50 estudiantes.

Las opciones son cualitativas, ya que se expresan con palabras.

- **Creamos una tabla para resumir los datos y establecer un criterio sobre su comportamiento:**

| Género musical | Frecuencia Absoluta |
|----------------|---------------------|
| Rock | 15 |
| Pop | 10 |
| Jazz | 8 |
| Electrónica | 10 |
| Clásica | 7 |

Interpretación:

La frecuencia absoluta nos muestra cuántas personas eligieron cada género musical.

Por ejemplo, 15 personas prefieren el Rock, 10 personas prefieren el Pop, etc.

Análisis:

Observamos que el género musical más popular entre los encuestados es el Rock.

Ejemplo de Distribución de Frecuencias para una Variable Ordinal: Nivel de Satisfacción

Supongamos que se ha encuestado a 100 clientes de un restaurante para evaluar su nivel de satisfacción con el servicio. Les pedimos que clasifiquen su satisfacción en una escala ordinal con las siguientes categorías:

- Muy insatisfecho
- Insatisfecho
- Neutral
- Satisfecho
- Muy satisfecho



- **Datos iniciales:**

Se han recopilado las respuestas de 100 clientes.

Las categorías son ordinales, ya que tienen un orden específico.

- **Tabla de distribución de frecuencias:**

Creamos una tabla para resumir los datos y establecer un criterio sobre su comportamiento:

| Nivel de Satisfacción | Frecuencia Absoluta |
|-----------------------|---------------------|
| Muy insatisfecho | 5 |
| Insatisfecho | 15 |
| Neutral | 30 |
| Satisfecho | 40 |
| Muy satisfecho | 10 |

Interpretación:

La frecuencia absoluta nos muestra cuántos clientes se encuentran en cada nivel de satisfacción.

Por ejemplo, 40 clientes están satisfechos, mientras que solo 5 clientes están muy insatisfechos.

Análisis:

Observamos que la mayoría de los clientes se encuentran en el nivel de satisfacción.

Ejemplo de Distribución de Frecuencias para una Variable Discreta: Temperaturas Diarias

Supongamos que durante el mes de junio cada día registramos la temperatura en grados Celsius y queremos analizar cómo se distribuyen estas temperaturas.



- **Datos iniciales:**

Hemos recopilado las temperaturas del mes de junio.

Las temperaturas son valores discretos, ya que se miden en grados enteros (por ejemplo, 25°C, 30°C, etc.).

- **Tabla de distribución de frecuencias:**

Creamos una tabla para resumir los datos y poder establecer un criterio sobre su comportamiento:

| Temperatura | Frecuencia Absoluta |
|-------------|---------------------|
| 15 °C | 6 |
| 20 °C | 7 |
| 25 °C | 10 |
| 30 °C | 4 |
| 35 °C | 3 |

Interpretación:

La frecuencia absoluta nos muestra cuántos días tuvieron cada temperatura.

Por ejemplo, 10 días tuvieron una temperatura de 25 °C, mientras que solo 3 días tuvieron una temperatura de 30°C.

Análisis:

Observamos que la mayoría de los días tuvieron temperaturas alrededor de 20 °C y 25 °C.

Hay menos días con temperaturas extremas (30 °C y 35 °C).

Tratamiento de variables Cuantitativas.

Medidas de tendencia central

Al describir grupos de observaciones, con frecuencia es conveniente resumir la información con un solo número. Este número que, para tal fin, suele situarse hacia el centro de la distribución de datos se denomina medida o parámetro de tendencia central o de centralización. Cuando se hace referencia únicamente a la posición de estos parámetros dentro de la distribución, independientemente de que ésta esté más o menos centrada, se habla de estas medidas como medidas de posición.



Las principales medidas de centralización que vamos a estudiar en este acápite son: media aritmética, mediana y moda.

Media Aritmética

Es el valor obtenido de la suma de todos los datos/valores dividida entre el número de datos sumados. Es sacar o realizar el promedio de los datos.



Su fórmula es:

$$\bar{x} = \frac{\sum x_i}{n}$$

Donde:

\bar{x} : Media aritmética o promedio

$\sum x_i$: Sumatoria de Datos

n: Número total de datos

Ejemplo 1: Número de programas semanales de cruceros al Caribe

Datos: 20, 73, 75, 80, 82.

Número total de datos: 5.

Media aritmética: 66

Solución:



$$\bar{x} = \frac{\sum x_i}{n}$$

$$\bar{x} = \frac{20 + 73 + 75 + 80 + 82}{5}$$

$$\bar{x} = \frac{330}{5}$$

$$\bar{x} = 66$$

Media aritmética: 66

Interpretación: En promedio, la agencia de cruceros recibe aproximadamente 66 programas semanales especiales de cruceros al Caribe.

Ejemplo 2: Ventas de botellas de agua durante la temporada de calor

Datos: 40, 55, 62, 43, 50, 60, 65.

Número total de datos: 7.

Solución:



$$\bar{x} = \frac{\sum x_i}{n}$$
$$\bar{x} = \frac{40 + 55 + 62 + 43 + 50 + 60 + 65}{7}$$
$$\bar{x} = \frac{375}{7}$$
$$\bar{x} = 53,57$$

Media aritmética: 53,57

Interpretación: Durante la temporada de calor, en promedio, se venden aproximadamente 53.57 botellas de 1 galón por hora en la tienda.

Mediana

La mediana m de un conjunto de n mediciones es el valor de x que cae en la posición media cuando las mediciones son ordenadas de menor a mayor.

Ejemplo 1: Edades de un grupo de personas:

Conjunto de datos: 20, 40, 30, 50, 25, 35, 45

Datos ordenados: 20, 25, 30, 35, 40, 45, 50

Número de datos: 7

Mediana: 35



Análisis e interpretación: En este conjunto de datos de edades, que consta de 7 observaciones, la mediana es 35. Esto significa que aproximadamente la mitad de las personas en el grupo tienen una edad de menos de 35 años y la otra mitad tienen más de 35 años.

Ejemplo 2: Ingresos mensuales de un grupo de empleados:

Conjunto de datos: \$2500, \$10000, \$2000, \$3500, \$3000

Datos ordenados: \$2000, \$2500, \$3000, \$3500, \$10000

Número de datos: 5

Mediana: \$3000



Análisis e interpretación: En este conjunto de datos de ingresos mensuales, que consta de 5 observaciones, la mediana es \$3000. Esto indica que aproximadamente la mitad de los empleados ganan menos de \$3000 al mes, y la otra mitad gana más de \$3000.

Ejemplo 3: Tiempos de entrega de un servicio de mensajería en días:

Conjunto de datos: 7, 4, 2, 10, 3, 12, 5

Datos ordenados: 2, 3, 4, 5, 7, 10, 12

Número de datos: 7

Mediana: 5 días



Análisis e interpretación: En este conjunto de datos de tiempos de entrega, que consta de 7 observaciones, la mediana de 5 días revela que la mitad de los paquetes se entregan en menos de 5 días, y la otra mitad se entrega más de 5 días.

Moda

La moda es la categoría que se presenta con más frecuencia o el valor de x que se presenta con más frecuencia. Cuando las mediciones en una variable continua se han agrupado como histograma de frecuencia o de frecuencia relativa, la clase con el pico más alto o frecuencia se llama clase modal, y el punto medio de esa clase se toma como la moda.



Ejemplo 1: Números de productos vendidos en una tienda

Conjunto de datos: detergente, mayonesa, harina, detergente, atún, mayonesa, detergente, harina, detergente

Número de datos: 9



Frecuencia de cada valor:

Mayonesa se repite 2 veces.

Harina se repite 2 veces.

Detergente se repite 4 veces.

Atún se repite 1 vez.

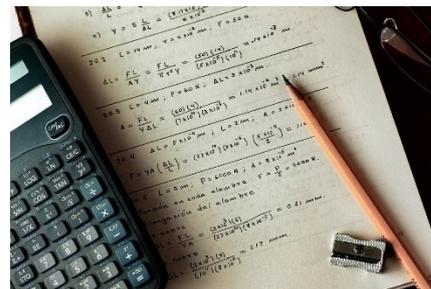
Moda: Detergente (se repite 4 veces)

Análisis e interpretación: En este conjunto de datos de números de productos vendidos, el valor 10 es la moda, ya que se repite 4 veces.

Ejemplo 2: Puntuaciones en un examen de matemáticas

Conjunto de datos: 85, 90, 75, 80, 90, 85, 90, 85, 80, 90

Número de datos: 10



Frecuencia de cada valor:

75 se repite 1 vez.

80 se repite 2 veces.

85 se repite 3 veces.

90 se repite 4 veces.

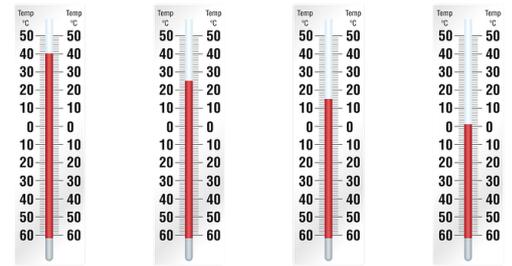
Moda: 90 (se repite 4 veces)

Análisis e interpretación: En este conjunto de datos de puntuaciones en un examen de matemáticas, la moda es 90, ya que es la puntuación más frecuente entre los estudiantes.

Ejemplo 3: Temperaturas registradas durante una semana

Conjunto de datos: 20°C, 22°C, 21°C, 20°C, 20°C, 22°C, 25°C

Número de datos: 7



Frecuencia de cada valor:

20°C se repite 3 veces.

21°C se repite 1 vez.

22°C se repite 2 veces.

25°C se repite 1 vez.

Moda: 20°C (se repite 3 veces)

Análisis e interpretación: En este conjunto de datos de temperaturas registradas durante una semana, la moda es 20°C, ya que fue la temperatura más frecuente durante ese período.

Técnicas de Visualización de datos

Diagrama de barras o columnas

Es un tipo de gráfico que constan de dos ejes, de los cuales se escoge a uno de ellos para representar a la variable de estudio de acuerdo a la distribución de frecuencias generada y el otro para representar la frecuencia de cada categoría, si es el eje vertical en el que se marcan las frecuencias el diagrama será de columnas, por el contrario, si es en el eje horizontal donde se representan las frecuencias, el diagrama será de barras.





Histograma

Un histograma de frecuencia relativa, para un conjunto de datos cuantitativo es una gráfica de barras en la que la altura de la barra muestra “con qué frecuencia” (medida como proporción o frecuencia relativa) las mediciones caen en una clase o subintervalo particular. Las clases o subintervalos se grafican a lo largo del eje horizontal.

¿Sabías que...?



Las variables ordinales, que permiten clasificar datos en un orden específico, han sido utilizadas desde tiempos antiguos. Los primeros censos romanos clasificaban a los ciudadanos en clases según su riqueza, lo que es un ejemplo temprano del uso de variables ordinales.



Curiosidades Estadísticas:



Los diagramas de barras y los histogramas no solo son útiles para los estadísticos sino también para los psicólogos. Estos gráficos pueden ayudar a ilustrar los resultados de estudios sobre comportamiento humano y percepción, facilitando la comprensión de patrones y tendencias entre diferentes grupos de personas.

| Clase 3 | |
|---|---|
| Asignatura | Matemática |
| Curso | 2do de BGU |
| Tema | Medidas de dispersión y visualización de Datos |
| Destreza   | <p>M.5.3.1. Calcular e interpretar la media, mediana, moda, rango, varianza y desviación estándar para datos no agrupados y agrupados, con apoyo de las TIC.</p> <p>M.5.3.2. Resolver y plantear problemas de aplicación de las medidas de tendencia central y de dispersión para datos agrupados, con apoyo de las TIC.</p> <p>M.5.3.3. Juzgar la validez de las soluciones obtenidas en los problemas de aplicación de las medidas de tendencia central y de dispersión para datos agrupados dentro del contexto del problema, con apoyo de las TIC.</p> |
| Indicador de evaluación | I.M.5.9.1. Calcula, con y sin apoyo de las TIC, las medidas de centralización y dispersión para datos agrupados y no agrupados; representa la información en gráficos estadísticos apropiados y los interpreta, juzgando su validez. (J.2.,I.3.) |

PARTE TEÓRICA

Medidas de dispersión de datos

Las medidas de variabilidad pueden ayudar a crear una imagen mental de la dispersión de los datos.

Varianza

Varianza de una población

La varianza de una población de N mediciones es el promedio de los cuadrados de las desviaciones de las mediciones alrededor de su media m . La varianza poblacional se denota con σ^2 y está dada por la fórmula:

$$\sigma^2 = \frac{\sum(x_i - \bar{x})^2}{N}$$

Donde:

σ^2 Es la varianza de la población.

x_i son los valores individuales en la población.

\bar{x} es la media de la población.

N es el tamaño de la población.

Ejemplos:

Ejemplo 1: Alturas de los árboles en un bosque

Supongamos que las alturas de los árboles en un bosque son: 10m, 12m, 15m, 11m, 13m. Calcularemos la varianza de esta población.



$$\sigma^2 = \frac{\sum(x_i - \bar{x})^2}{N}$$

Paso 1: Calculamos la media

$$\bar{x} = \frac{\sum x_i}{n}$$

$$\bar{x} = \frac{10 + 12 + 15 + 11 + 13}{5}$$

$$\bar{x} = \frac{61}{5}$$

$$\bar{x} = 12,2$$

Paso 2: Calculamos la varianza

$$\sigma^2 = \frac{(10 - 12,2)^2 + (12 - 12,2)^2 + (15 - 12,2)^2 + (11 - 12,2)^2 + (13 - 12,2)^2}{5}$$

$$\sigma^2 = \frac{(-2,2)^2 + (-0,2)^2 + (2,8)^2 + (-1,2)^2 + (0,8)^2}{5}$$

$$\sigma^2 = \frac{4,84 + 0,04 + 7,84 + 1,44 + 0,64}{5}$$

$$\sigma^2 = \frac{14,8}{5}$$

$$\sigma^2 = 2,96$$

Análisis e Interpretación:

La varianza calculada es 2.96. Esto significa que las alturas de los árboles en el bosque varían en promedio alrededor de 2.96 metros con respecto a la media.

Una varianza más alta indica una mayor dispersión en las alturas de los árboles.

Ejemplo 2: Edades de todos los estudiantes en una escuela

Supongamos que las edades de todos los estudiantes en una escuela son: 10 años, 12 años, 11 años, 13 años, 14 años. Calcularemos la varianza de esta población.



$$\sigma^2 = \frac{\sum(x_i - \bar{x})^2}{N}$$

Paso 1: Calculamos la media

$$\bar{x} = \frac{\sum x_i}{n}$$

$$\bar{x} = \frac{10 + 12 + 11 + 13 + 14}{5}$$

$$\bar{x} = \frac{60}{5}$$

$$\bar{x} = 12$$

Paso 2: Calculamos la varianza

$$\sigma^2 = \frac{(10 - 12)^2 + (12 - 12)^2 + (11 - 12)^2 + (13 - 12)^2 + (14 - 12)^2}{5}$$

$$\sigma^2 = \frac{(-2)^2 + (0)^2 + (-1)^2 + (1)^2 + (2)^2}{5}$$

$$\sigma^2 = \frac{4 + 0 + 1 + 1 + 4}{5}$$

$$\sigma^2 = \frac{10}{5}$$

$$\sigma^2 = 2$$

Análisis e Interpretación:

La varianza calculada es 2. Esto significa que las edades de los estudiantes en la escuela varían en promedio alrededor de 2 años con respecto a la media.

Una varianza más alta indica una mayor dispersión en las edades de los estudiantes.

Varianza de una muestra

La varianza de una muestra de n mediciones es la suma de las desviaciones cuadradas de las mediciones alrededor la media \bar{x} dividida entre $(n-1)$. La varianza muestral se denota con s^2 y está dada por la fórmula

$$s^2 = \frac{\sum(x_i - \bar{x})^2}{n - 1}$$

Donde:

s^2 es la varianza de la muestra.

x_i son los valores individuales en la población.

\bar{x} es la media de la muestra.

n es el tamaño de la muestra.

Ejemplos

Ejemplo 1: Calificaciones de un grupo de estudiantes

Supongamos que las calificaciones de un grupo de estudiantes en un examen son: 80, 85, 90, 75, 95. Calcularemos la varianza de esta muestra.



$$s^2 = \frac{\sum(x_i - \bar{x})^2}{n - 1}$$

Paso 1: Calculamos la media

$$\bar{x} = \frac{\sum x_i}{n}$$

$$\bar{x} = \frac{80 + 85 + 90 + 75 + 95}{5}$$

$$\bar{x} = \frac{425}{5}$$

$$\bar{x} = 85$$

Paso 2: Calculamos la varianza

$$s^2 = \frac{(80 - 85)^2 + (85 - 85)^2 + (90 - 85)^2 + (75 - 85)^2 + (95 - 85)^2}{5 - 1}$$

$$s^2 = \frac{(-5)^2 + (0)^2 + (5)^2 + (-10)^2 + (10)^2}{4}$$

$$s^2 = \frac{25 + 0 + 25 + 100 + 100}{4}$$

$$s^2 = \frac{250}{4}$$

$$s^2 = 62,5$$

Análisis e Interpretación:

La varianza calculada es 62,5. Esto significa que las calificaciones de los estudiantes en la muestra varían en promedio alrededor de 62,5 puntos con respecto a la media.

Una varianza más alta indica una mayor dispersión en las calificaciones.

Ejemplo 2: Ingresos de una muestra de hogares en una ciudad

Supongamos que los ingresos mensuales (en miles de dólares) de una muestra de hogares en una ciudad son: 4, 5, 6, 4,5, 5,5. Calcularemos la varianza de esta muestra.



$$s^2 = \frac{\sum(x_i - \bar{x})^2}{n - 1}$$

Paso 1: Calculamos la media

$$\bar{x} = \frac{\sum x_i}{n}$$

$$\bar{x} = \frac{4 + 5 + 6 + 4,5 + 5,5}{5}$$

$$\bar{x} = \frac{25}{5}$$

$$\bar{x} = 5$$

Paso 2: Calculamos la varianza

$$s^2 = \frac{(4 - 5)^2 + (5 - 5)^2 + (6 - 5)^2 + (4,5 - 5)^2 + (5,5 - 5)^2}{5 - 1}$$

$$s^2 = \frac{(-1)^2 + (0)^2 + (1)^2 + (-0,5)^2 + (0,5)^2}{4}$$

$$s^2 = \frac{1 + 0 + 1 + 0,25 + 0,25}{4}$$

$$s^2 = \frac{2,50}{4}$$

$$s^2 = 0,625$$

Análisis e Interpretación:

La varianza calculada es 0.625. Esto significa que los ingresos mensuales de los hogares en la muestra varían en promedio alrededor de \$625 con respecto a la media (medidos en miles de dólares).

Una varianza más baja indica una menor dispersión en los ingresos.

Desviación estándar

La desviación estándar de un conjunto de mediciones es igual a la raíz cuadrada positiva de la varianza.

Mención sobre datos reales

En nuestra vida cotidiana, interactuamos con conjuntos masivos de datos sin siquiera darnos cuenta. Desde las transacciones bancarias hasta las redes sociales, desde los sensores en nuestros dispositivos hasta los registros médicos, todo genera datos.



Los datos reales son registros auténticos y representan la realidad, por ejemplo: los patrones económicos, la propagación de enfermedades, las preferencias de los consumidores y mucho más. Estos datos no solo son números; son ventanas que nos permiten comprender a profundidad nuestro mundo.

Cuando los científicos, los economistas o los expertos en marketing se sumergen en estos datos pueden llegar a predecir tendencias, tomar decisiones informadas y resolver problemas complejos. Pero, ¿cómo se trabaja con estos datos? ¿Qué herramientas y técnicas se utilizan?



Clase 4

| | |
|--|--|
| Asignatura | Matemática |
| Curso | 2do de BGU |
| Tema | Introducción a datos abiertos y Python |
| Destreza  | M.5.3.1. Calcular e interpretar la media, mediana, moda, rango, varianza y desviación estándar para datos no agrupados y agrupados, con apoyo de las TIC. M.5.3.2. Resolver y plantear problemas de aplicación de las medidas de tendencia central y de dispersión para datos agrupados, con apoyo de las TIC. M.5.3.3. Juzgar la validez de las soluciones obtenidas en los problemas de aplicación de las medidas de tendencia central y de dispersión para datos agrupados dentro del contexto del problema, con apoyo de las TIC. |
| Indicador de evaluación | I.M.5.9.1. Calcula, con y sin apoyo de las TIC, las medidas de centralización y dispersión para datos agrupados y no agrupados; representa la información en gráficos estadísticos apropiados y los interpreta, juzgando su validez. (J.2.,I.3.) |

PARTE TEÓRICA

Datos abiertos

Los datos abiertos se refieren a conjuntos de datos que están disponibles para que cualquier persona los acceda, use y comparta, sin restricciones, patentes u otros mecanismos de control. Representan una fuente invaluable de información que abarca una amplia gama de temas, desde el clima y el transporte hasta la salud y la educación.

La importancia de los datos abiertos radica en su capacidad para fomentar la transparencia, la innovación y la participación ciudadana. Al hacer que la información sea accesible para todos, los datos abiertos promueven la rendición de cuentas y permiten que los ciudadanos tomen decisiones informadas sobre temas que afectan sus vidas. Además, proporcionan un recurso invaluable para investigadores, empresas y organizaciones sin fines de lucro que buscan abordar desafíos sociales, económicos y ambientales.



Sin embargo, trabajar con estas vastas cantidades de datos puede resultar abrumador y desafiante para los métodos tradicionales de análisis. Los procesos manuales simplemente no son viables cuando se trata de manipular grandes cantidades de información. Es en este punto donde entra en juego el papel fundamental del uso de herramientas tecnológicas como, por ejemplo: excel, R o Python, entre otras. En este caso, Python representa una alternativa esencial para trabajar con este tipo de datos, ya que es un software que ofrece una amplia gama de herramientas y bibliotecas especializadas en el manejo y análisis de grandes conjuntos de datos.



¿Qué es Python?

Python es un lenguaje de programación con una sintaxis elegante que permite escribir programas de una manera fácil, ofrece un entorno interactivo el cual detecta errores de programación y proporciona información de utilidad para su correcta ejecución, además de contar con una amplia variedad de estructuras de datos de fácil manejo.



Como se instala Python.

Para hacer uso de Python debemos hacer lo siguiente:

Descargar Anaconda: Visitar el sitio web oficial de Anaconda (<https://www.anaconda.com/>) y descargar la versión adecuada para tu sistema operativo (Windows, macOS o Linux).

Instalación de Anaconda: Una vez que se haya completado la descarga, seguir las instrucciones de instalación proporcionadas en el sitio web de Anaconda para tu sistema operativo.

Finalización de la instalación: Una vez completada la instalación, se puede verificar si Anaconda se haya instalado correctamente abriendo una terminal o símbolo del sistema y escribiendo `conda --version`. Se debería ver la versión de Anaconda instalada.

Jupyter Notebook: Después de instalar Anaconda, se puede instalar Jupyter Notebook, el cual es una aplicación web de código



Videos complementarios:
Cómo instalar Python con ANACONDA   :
<https://www.youtube.com/watch?v=OmmkLYIRGzo>
Descargar e instalar ANACONDA para Python:
https://www.youtube.com/watch?v=0lDixuz_K88

abierto que permite crear y compartir documentos interactivos que contienen código, visualizaciones, etc. Una vez que Jupyter Notebook esté instalado y abierto en el navegador, se puede crear un nuevo cuaderno haciendo clic en el botón "New" en la esquina superior derecha y seleccionando el kernel de Python que se desea utilizar (Python 2 o Python 3)

¿Sabías que...?



Durante la pandemia de COVID-19, los datos abiertos jugaron un papel crucial al permitir a los investigadores y al público en general acceder a información actualizada sobre la propagación del virus, las tasas de vacunación y las medidas de respuesta.



Curiosidades Estadísticas:



Ecuador, hogar de una rica biodiversidad, utiliza datos abiertos para la conservación del medio ambiente. Los investigadores tienen acceso a datos sobre especies en peligro, áreas protegidas y cambios en el uso de la tierra, lo que ayuda a desarrollar estrategias efectivas para la conservación.

| Clase 5 | |
|---|---|
| Asignatura | Matemática |
| Curso | 2do de BGU |
| Tema | Proyecto guiado de análisis de datos con Python (Anexo 1.) |
| Destreza   | <p>M.5.3.1. Calcular e interpretar la media, mediana, moda, rango, varianza y desviación estándar para datos no agrupados y agrupados, con apoyo de las TIC.</p> <p>M.5.3.2. Resolver y plantear problemas de aplicación de las medidas de tendencia central y de dispersión para datos agrupados, con apoyo de las TIC.</p> <p>M.5.3.3. Juzgar la validez de las soluciones obtenidas en los problemas de aplicación de las medidas de tendencia central y de dispersión para datos agrupados dentro del contexto del problema, con apoyo de las TIC.</p> |
| Indicador de evaluación | I.M.5.9.1. Calcula, con y sin apoyo de las TIC, las medidas de centralización y dispersión para datos agrupados y no agrupados; representa la información en gráficos estadísticos apropiados y los interpreta, juzgando su validez. (J.2.,I.3.) |
| Metodología | Aprendizaje basado en proyectos: Este proyecto se basa en el enfoque pedagógico del aprendizaje basado en proyectos (ABP). En lugar de enseñarles conceptos de manera aislada, los estudiantes aprenderán a través de la experiencia de trabajar en un proyecto práctico centrado en el análisis y visualización de datos de vehículos nuevos. |

PARTE PRÁCTICA

Exploración y Visualización de Datos

El proyecto "Exploración y Visualización de Datos" tiene como objetivo enseñar a los estudiantes a cargar, limpiar y analizar datos utilizando Python, una herramienta fundamental en el análisis de datos. A través de este proyecto, los estudiantes aprenderán a utilizar librerías de Python para manipular datos, calcular medidas estadísticas y crear visualizaciones informativas. Este conocimiento es crucial en el mundo actual, donde el análisis de grandes volúmenes de datos es esencial para la toma de decisiones informadas.



Objetivos de aprendizaje: Espero que al final de este proyecto, todos ustedes sean capaces de manejar y analizar datos utilizando Python, entender y calcular medidas de tendencia central y de dispersión, y visualizar datos a través de histogramas y gráficos de barras. Además, aprenderán a interpretar y analizar datos en un contexto real.

Etapas del Análisis de Datos

Importación de Librerías:

- En esta etapa, aprenderemos a importar las librerías necesarias para nuestro análisis. Las librerías son paquetes de código que nos permiten realizar tareas específicas de manera más eficiente:

```
import pandas as pd
import matplotlib.pyplot as plt
import numpy as np
import seaborn as sns
```

| Código | Explicación |
|---------------------------------|--|
| import pandas as pd | Manipulación y análisis de datos con DataFrames. |
| import matplotlib.pyplot as plt | Visualización de datos en 2D, creación de gráficos. |
| import numpy as np | Soporte para matrices grandes y funciones matemáticas. |
| import seaborn as sns | Visualización de datos estadísticos con gráficos atractivos. |

Importación de Datos:

- Cargaremos un archivo Excel que contiene los datos de vehículos nuevos en 2022, descargado de la plataforma de Datos Abiertos de Ecuador.

```
from google.colab import files

uploaded = files.upload()
```

```
df=pd.read_excel("SRI_Vehiculos_Nuevos_2022.xlsx")
```

```
df
```

| Código | Explicación |
|--------------------------------|---|
| from google.colab import files | Importa el módulo `files` de Google Colab para manejar archivos. |
| uploaded = files.upload() | Abre un cuadro para cargar archivos desde la computadora a Colab. |
| df = pd.read_excel("Archivo") | Lee el archivo Excel y lo carga en un DataFrame `df`. |
| df | Muestra las primeras filas del DataFrame `df`. |

- Dado que los nombres de las cabeceras son muy largos e incluso tienen caracteres poco amigables, las vamos a cambiar:

```
df.columns = ["Categoria", "C.vehiculo", "T.transaccion", "Marca", "Modelo", "Pais",
              "Añomodelo", "Clase", "Sub_clase", "Tipo", "Avaluo", "Fecha_proceso",
              "Tipo_servicio", "Cilindraje", "Tipo_combustible", "Fecha_compra",
              "Canton", "Color_1", "Color_2", "Persona_natural"]
```

```
df
```

| Código | Explicación |
|---------------------------------|---|
| <code>df.columns = [...]</code> | Cambia los nombres de las columnas del DataFrame `df` a nombres más descriptivos. |
| <code>df</code> | Muestra las primeras filas del DataFrame `df`. |

Limpieza y Preparación de Datos:

- Eliminaremos datos duplicados para asegurar que nuestro análisis sea preciso

```
df = df.drop_duplicates()
```

```
df
```

| Código | Explicación |
|--|--|
| <code>df = df.drop_duplicates()</code> | Elimina filas duplicadas del DataFrame `df`. |
| <code>df</code> | Muestra las primeras filas del DataFrame `df`. |

Medidas de Tendencia Central:

Calcularemos la media, mediana y moda del cilindraje de los vehículos.

Media:

Problema: Un fabricante de automóviles está interesado en el cilindraje promedio de los vehículos vendidos en 2022 para planificar la producción del próximo año.



- Para ello utilizaremos el siguiente código:

```
media_cilindraje = df['Cilindraje'].mean()
print('La media del cilindraje es: {}'.format(media_cilindraje))
```

| Código | Explicación |
|---|--|
| <code>media_cilindraje = df['Cilindraje'].mean()</code> | Calcula la media del cilindraje en <code>df</code> . |
| <code>print('La media del cilindraje es: {}'.format(media_cilindraje))</code> | Imprime el cilindraje promedio con <code>format</code> . |

Mediana:



Problema: Un importador de vehículos desea traer al país los modelos más comunes. Para ello, es esencial conocer el cilindraje mediano de los vehículos vendidos.

- Para ello utilizaremos el siguiente código:

```
mediana_cilindraje = df['Cilindraje'].median()
print('La mediana del cilindraje es: {}'.format(mediana_cilindraje))
```

| Código | Explicación |
|--|---|
| <pre>mediana_cilindraje = df['cilindraje'].median() print('La mediana del cilindraje es: {}'.format(mediana_cilindraje))</pre> | Calcula la mediana del cilindraje en <code>df</code> . Imprime la mediana del cilindraje con <code>format</code> . |

Moda:

Problema: Una empresa de combustibles está buscando ajustar su producción según el tipo de combustible más utilizado en los vehículos vendidos en 2022.



- Para ello utilizaremos el siguiente código:

```
moda_tipo_combustible = df['Tipo_combustible'].mode()[0]
print('El tipo de combustible más común es: {}'.format(modato_tipo_combustible))
```

| Código | Explicación |
|--|--|
| <pre>moda_tipo_combustible = df['Tipo_combustible'].mode()[0] print('El tipo de combustible más común es: {}'.format(modato_tipo_combustible))</pre> | Calcula la moda del tipo de combustible en <code>df</code> . Imprime el tipo de combustible más frecuente con <code>form</code> . |

Medidas de Dispersión:

Calcularemos la varianza y la desviación estándar del cilindraje de los vehículos.

Varianza y Desviación Estándar:



Problema: Un analista de mercado está buscando entender la variabilidad del cilindraje de los vehículos vendidos en 2022.

- Para ello utilizaremos el siguiente código:

```
varianza_cilindraje = df['Cilindraje'].var()
desviacion_cilindraje = df['Cilindraje'].std()
print('La varianza del cilindraje es: {}'.format(varianza_cilindraje))
print('La Desviacion estandar del cilindraje es: {}'.format(desviacion_cilindraje))
```

| Código | Explicación |
|--|---|
| <code>varianza_cilindraje = df['cilindraje'].var()</code> | Calcula la varianza del cilindraje en <code>df</code> . |
| <code>desviacion_cilindraje = df['cilindraje'].std()</code> | Calcula la desviación estándar del cilindraje en <code>df</code> . |
| <code>print('La varianza del cilindraje es: {}'.format(varianza_cilindraje))</code> | Imprime la varianza del cilindraje con <code>format</code> . |
| <code>print('La desviación estándar del cilindraje es: {}'.format(desviacion_cilindraje))</code> | Imprime la desviación estándar del cilindraje con <code>format</code> . |

Visualización de Datos:

Crearemos un gráfico de barras para visualizar las clases de vehículos más vendidas en 2022.

Gráfico de Barras:

Problema: Un concesionario de automóviles está interesado en conocer las clases de vehículos más vendidas en 2022 para ajustar su inventario.



- Para ello utilizaremos el siguiente código:

```
frecuencia = df['Clase'].value_counts()
print(frecuencia)
```

```
# Creamos el gráfico de barras
plt.figure(figsize=(10,6)) # Tamaño del gráfico
plt.bar(frecuencia.index, frecuencia.values, color='skyblue', edgecolor='black')
plt.title('Gráfico de Barras de las Clases de Vehículos')
plt.xlabel('Clase')
plt.ylabel('Número de Vehículos')
plt.xticks(rotation=90) # Rotamos las etiquetas del eje x
plt.show()
```

| Código | Explicación |
|--|---|
| <code>frecuencia_clase = df['Clase'].value_counts()</code> | Cuenta la frecuencia de cada clase en <code>df</code> . |
| <code>print(frecuencia_clase)</code> | Imprime la frecuencia de cada clase. |
| <code>plt.figure(figsize=(10,6))</code> | Crea una figura de 10x6 para el gráfico. |
| <code>plt.bar(frecuencia_clase.index, frecuencia_clase.values, color='skyblue', edgecolor='black')</code> | Crea un gráfico de barras con clases y frecuencias. |
| <code>plt.title('Gráfico de Barras de las Clases de Vehículos')</code> <code>plt.xlabel('Clase')</code> <code>plt.ylabel('Número de Vehículos')</code> | Añade título y etiquetas a los ejes del gráfico. |
| <code>plt.xticks(rotation=90)</code> | Rota las etiquetas del eje x 90 grados. |
| <code>plt.show()</code> | Muestra el gráfico. |

| Clase 6 | |
|---|---|
| Asignatura | Matemática |
| Curso | 2do de BGU |
| Tema | Proyecto conjunto de análisis de datos con Python (Anexo 2.) |
| Destreza   | <p>M.5.3.1. Calcular e interpretar la media, mediana, moda, rango, varianza y desviación estándar para datos no agrupados y agrupados, con apoyo de las TIC.</p> <p>M.5.3.2. Resolver y plantear problemas de aplicación de las medidas de tendencia central y de dispersión para datos agrupados, con apoyo de las TIC.</p> <p>M.5.3.3. Juzgar la validez de las soluciones obtenidas en los problemas de aplicación de las medidas de tendencia central y de dispersión para datos agrupados dentro del contexto del problema, con apoyo de las TIC.</p> |
| Indicador de evaluación | I.M.5.9.1. Calcula, con y sin apoyo de las TIC, las medidas de centralización y dispersión para datos agrupados y no agrupados; representa la información en gráficos estadísticos apropiados y los interpreta, juzgando su validez. (J.2.,I.3.) |
| Metodología | Aprendizaje basado en proyectos: Este proyecto se basa en el enfoque pedagógico del aprendizaje basado en proyectos (ABP). En lugar de enseñarles conceptos de manera aislada, los estudiantes aprenderán a través de la experiencia de trabajar en un proyecto práctico centrado en el análisis y visualización de datos de altura horaria predicha. |

PARTE PRÁCTICA

Exploración y Visualización de Datos de Mareas

El proyecto "Exploración y Visualización de Datos de Mareas" tiene como objetivo utilizar Python y Google Colab para analizar datos reales sobre las mareas en varias playas a lo largo del año 2024. A través de este proyecto, aprenderemos a importar datos, calcular medidas estadísticas como la media, mediana, moda, varianza y desviación estándar, y crear visualizaciones informativas como histogramas y gráficos de barras. Este conocimiento es esencial para comprender y tomar decisiones informadas sobre los patrones de mareas, un fenómeno crucial para las comunidades costeras.



Objetivo de Aprendizaje: Al finalizar este proyecto, se espera que puedas adquirir las habilidades necesarias para importar, analizar y visualizar datos de mareas utilizando Python y Google Colab. Aprenderás a calcular medidas estadísticas clave como la media, mediana, moda, varianza y desviación estándar, así como a crear visualizaciones informativas como histogramas y gráficos de barras. Estas

habilidades te permitirán comprender patrones de mareas y tomar decisiones informadas basadas en datos reales.

Etapas del Análisis de Datos de Mareas

Importación de Librerías:

- En esta etapa, aprenderemos a importar las librerías necesarias para nuestro análisis. Las librerías son paquetes de código que nos permiten realizar tareas específicas de manera más eficiente:

```
import pandas as pd
import matplotlib.pyplot as plt
import numpy as np
import seaborn as sns
import datetime
```

| Código | Explicación |
|-------------------|---|
| pandas | Manipula y analiza datos en DataFrames, facilita filtrado, selección y cálculos estadísticos. |
| matplotlib.pyplot | Crea gráficos y visualizaciones, esencial para representar datos. |
| numpy | Soporta arreglos multidimensionales y operaciones matemáticas avanzadas. |
| seaborn | Ofrece estilos estéticos y funciones para visualizaciones estadísticas avanzadas. |
| datetime | Maneja y formatea fechas y horas fácilmente. |

Importación de Datos:

- Ahora procederemos a cargar nuestra base de datos de mareas para comenzar nuestro análisis.

```
from google.colab import files
uploaded = files.upload()
```

```
df = pd.read_excel("inocar_altura_horaria_predicha_2023diciembre.xls")
```

```
df
```

| Código | Explicación |
|--|---|
| from google.colab import files | Importa el módulo <code>files</code> de Google Colab para manejar archivos. |
| uploaded = files.upload() | Abre una ventana para cargar archivos desde la computadora a Colab. |
| df = pd.read_excel("nombre del archivo") | Lee el archivo Excel y carga los datos en un DataFrame <code>df</code> . |
| df | Muestra las primeras filas del DataFrame para verificar la carga. |

Preparación de Datos:

- Aseguraremos que nuestros datos estén limpios y listos para el análisis.

Cambio de nombres de columnas

```
df.columns = ["FECHA", "CRUZ", "LORENZO", "ESMERALDAS",
              "LIBERTAD", "GUAYAS", "CARAQUEZ", "MANTA", "BOLIVAR", "BALTRA"]
```

```
df
```

| Código | Explicación |
|---------------------------------|--|
| <code>df.columns = [...]</code> | Cambia los nombres de las columnas en el DataFrame <code>df</code> . |
| <code>df</code> | Muestra las primeras filas del DataFrame <code>df</code> . |

Medidas de Tendencia Central:

Media

Uno de nuestros objetivos iniciales es calcular la altura de marea promedio para cada playa. Este valor promedio nos dará una idea de la altura típica de las mareas en cada ubicación y nos permitirá comparar entre diferentes playas.



- **Cálculo de la media:** Recordemos que la media es la suma de todos los valores divididos por el número total de valores. En este caso, nos va a servir para entender cuál es la altura de marea ‘promedio’ en cada playa.
- **Para ello vamos a utilizar el siguiente código:**

```
nombre_columna_playa = input("Ingresa el nombre de la columna de la playa: ")
# Filtrar los datos para la playa seleccionada
playa_seleccionada = df[nombre_columna_playa]
# Calcular la media:
media = playa_seleccionada.mean()
print("Media:", media)
```

| Código | Explicación |
|--|--|
| <code>nombre_columna_playa = input(...)</code> | Solicita al usuario el nombre de la columna de la playa. |
| <code>playa_seleccionada = df[nombre_columna_playa]</code> | Filtra el DataFrame <code>df</code> según la columna seleccionada. |
| <code>media = playa_seleccionada.mean()</code> | Calcula la media de los datos de la playa seleccionada. |
| <code>print("Media:", media)</code> | Imprime la media de la altura de la marea en la playa. |

Mediana



El siguiente objetivo es calcular la mediana de las alturas de marea para cada playa. La mediana nos proporcionará un valor central que separa el conjunto de datos ordenados en dos partes iguales, lo que nos ayudará a entender mejor la distribución de las alturas de marea.

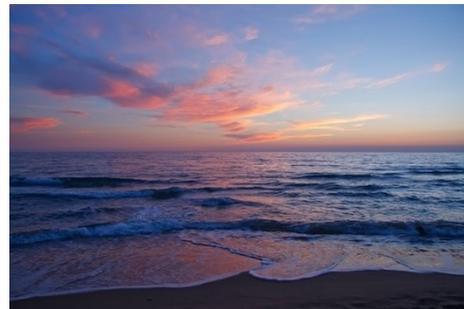
- **Cálculo de la mediana:** Otra medida que podemos utilizar para entender nuestros datos es la mediana. A diferencia de la media, la mediana es el valor que se encuentra justo en el medio de nuestros datos cuando están ordenados de menor a mayor. Recordemos que la mediana nos da una idea del valor ‘central’ de nuestros datos. En este caso, nos va a servir para entender cuál es la altura de marea ‘típica’ en cada playa, sin que se vea afectada por valores extremadamente altos o bajos.
- **Para ello vamos a utilizar el siguiente código:**

```
nombre_columna_playa = input("Ingresa el nombre de la columna de la playa: ")
# Filtrar los datos para la playa seleccionada
playa_seleccionada = df[nombre_columna_playa]
# Calcular la mediana:
mediana = playa_seleccionada.median()
print("Mediana:", mediana)
```

| Código | Explicación |
|--|--|
| <code>nombre_columna_playa = input(...)</code> | Solicita al usuario el nombre de la columna de la playa. |
| <code>playa_seleccionada = df[nombre_columna_playa]</code> | Filtra el DataFrame <code>df</code> según la columna seleccionada. |
| <code>mediana = playa_seleccionada.median()</code> | Calcula la mediana de los datos de la playa seleccionada. |
| <code>print("Mediana:", mediana)</code> | Imprime la mediana de la altura de la marea en la playa. |

Moda

Nuestro objetivo es calcular la moda de las alturas de marea para cada playa. La moda nos proporcionará el valor que aparece con mayor frecuencia en el conjunto de datos, lo que nos ayudará a identificar los valores más comunes de alturas de marea.



- **Cálculo de la moda:** Finalmente, podemos calcular la moda, que es el valor que aparece con más frecuencia en nuestros datos. Recordemos que la moda nos da una idea de los valores ‘comunes’ en nuestros datos. En este caso, nos va a servir para entender cuál es la altura de marea más frecuente en cada playa.
- **Para ello vamos a utilizar el siguiente código:**

```
nombre_columna_playa = input("Ingresa el nombre de la columna de la playa: ")
# Filtrar los datos para la playa seleccionada
playa_seleccionada = df[nombre_columna_playa]
# Calcular la moda:
moda = playa_seleccionada.mode().iloc[0]
print("Moda:", moda)
```

| Código | Explicación |
|--|--|
| <code>nombre_columna_playa = input(...)</code> | Solicita al usuario el nombre de la columna de la playa. |
| <code>playa_seleccionada = df[nombre_columna_playa]</code> | Filtra el DataFrame <code>df</code> según la columna seleccionada. |
| <code>moda = playa_seleccionada.mode().iloc[0]</code> | Calcula la moda de los datos de la playa seleccionada. |
| <code>print("Moda:", moda)</code> | Imprime la moda de la altura de marea en la playa. |

Medidas de dispersión de datos

Problema general: Ahora que hemos entendido qué es ‘típico’ en nuestros datos de mareas mediante las medidas de tendencia central, el siguiente paso es entender cómo se ‘dispersan’ o varían estos datos

Desviación estándar



Nuestro objetivo será calcular la desviación estándar de las alturas de marea para cada playa. La desviación estándar nos proporcionará una medida de cuánto varían las alturas de marea con respecto a su media, lo que nos ayudará a entender la dispersión de los datos y la consistencia de las mareas en cada ubicación.

- **Cálculo de la desviación estándar:** La desviación estándar es otra medida de dispersión, que es simplemente la raíz cuadrada de la varianza. A diferencia de la varianza, la desviación estándar está en las mismas unidades que los datos, lo que a menudo la hace más fácil de interpretar. Vamos a calcular la desviación estándar de las mareas para cada playa con el siguiente código:

```
nombre_columna_playa = input("Ingresa el nombre de la columna de la playa: ")
# Filtrar los datos para la playa seleccionada
playa_seleccionada = df[nombre_columna_playa]
# Calcular la desviación estándar
desviacion_estandar = playa_seleccionada.std()
print("Desviación estándar: ", desviacion_estandar)
```

| Código | Explicación |
|---|--|
| <code>nombre_columna_playa = input(...)</code> | Solicita al usuario el nombre de la columna de la playa. |
| <code>playa_seleccionada = df[nombre_columna_playa]</code> | Filtra el DataFrame <code>df</code> según la columna seleccionada. |
| <code>desviacion_estandar = playa_seleccionada.std()</code> | Calcula la desviación estándar de los datos de la playa. |
| <code>print("Desviación Estándar:", desviacion_estandar)</code> | Imprime la desviación estándar de la altura de marea en la playa. |

Técnicas de visualización de datos

Problema general: Ahora que hemos calculado las medidas de tendencia central y de dispersión para nuestras mareas, el siguiente paso es visualizar estos datos. ¿Cómo podemos representar gráficamente nuestras mareas para entender mejor su comportamiento? Aquí es donde entran en juego las técnicas visuales de datos: los histogramas y los gráficos de barras.

Nuestro objetivo es crear un histograma que muestre la distribución de alturas de marea para una de las playas seleccionadas. Un histograma es una representación visual de la distribución de frecuencia de los datos, donde las alturas de marea se agrupan en intervalos y se muestra la frecuencia de ocurrencia de cada intervalo.



- **Creación de histogramas:** Un histograma te mostrará cuántas mareas caen en cada rango de alturas. Para ello utilizaremos el siguiente código:

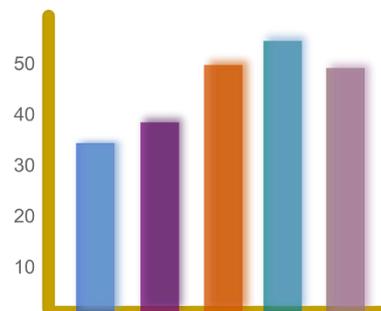
```
nombre_playa = input("Ingresa el nombre de la columna playa: ")
mes= input("Ingresa el mes (en formato numérico): ")

df['FECHA'] = pd.to_datetime(df['FECHA'])

# Filtrar los datos para el mes seleccionado
playa_seleccionada = df[df['FECHA'].dt.month == int(mes)][nombre_playa]

# Crear un histograma
playa_seleccionada.hist(bins=50)
plt.title(f"Histograma de las mareas para {nombre_playa} en el mes {mes}")
plt.show()
```

Creación de gráficos de barras: Los gráficos de barras son otra herramienta útil para visualizar nuestros datos. Pero, ¿qué es realmente un gráfico de barras? Imagina que quieres comparar la altura media de las mareas entre diferentes playas. Un gráfico de barras te mostrará esta comparación de manera clara y fácil de entender. Vamos a crear un gráfico de barras de la altura media de las mareas para cada playa con el siguiente código:



```

nombre_playa = input("Ingresa el nombre de la columna de la playa: ")
df['FECHA'] = pd.to_datetime(df['FECHA'])
# Filtrar los datos para la playa seleccionada
playa_seleccionada = df[nombre_playa]
# Crear un gráfico de barras
playa_seleccionada.groupby(df['FECHA'].dt.month).mean().plot(kind='bar')
plt.title(f"Alturas promedio de las mareas para {nombre_playa} por mes")
plt.xlabel('Mes del año')
plt.ylabel('Altura promedio de la marea')
plt.show()

```

| Código | Explicación |
|---|---|
| nombre_playa = input("Ingresa el nombre de la columna playa: ") | Solicita al usuario el nombre de la columna de la playa. |
| mes = input("Ingresa el mes (Ejemplo: Julio=7): ") | Solicita al usuario el número del mes para analizar. |
| df['FECHA'] = pd.to_datetime(df['FECHA']) | Convierte la columna 'FECHA' a formato datetime. |
| playa_seleccionada = df[df['FECHA'].dt.month == int(mes)][nombre_playa] | Filtra los datos para el mes y la columna de la playa. |
| playa_seleccionada.hist(bins=50) | Crea un histograma con 50 bins para las alturas de marea. |
| plt.title(f"Histograma de las mareas para {nombre_playa} en el mes {mes}") | Establece el título del histograma con el nombre de la playa y el mes. |
| plt.show() | Muestra el histograma. |

RESULTADOS ESPERADOS

Con la implementación de esta propuesta didáctica, se espera crear un ambiente de aprendizaje dinámico y motivador para los estudiantes de 2do de Bachillerato General Unificado (BGU). Los estudiantes no solo adquirirán una comprensión profunda de los conceptos fundamentales de la Estadística Descriptiva, sino que también desarrollarán habilidades prácticas en el uso de Python para el análisis y visualización de datos. Al trabajar con datos reales y aplicar sus conocimientos teóricos a problemas prácticos, se espera que los estudiantes desarrollen un pensamiento crítico y reflexivo, así como una capacidad para resolver problemas de manera efectiva.

Además, se anticipa que los estudiantes alcanzarán las destrezas planteadas en cada etapa del curso, utilizando medios digitales y estrategias comunicacionales. Se fomentará el trabajo colaborativo y la capacidad de comunicar resultados de manera clara y efectiva. La experiencia de trabajar con datos abiertos y proyectos basados en situaciones reales preparará a los estudiantes para futuros estudios y carreras en campos donde el análisis de datos es fundamental, asegurando que estén bien equipados para enfrentar los desafíos del siglo XXI.

BIBLIOGRAFIA

- Marzal, A., Gracia, I., y García, P. (2014). *Introducción a la programación con Python 3*. Universitat Jaume I. <https://doi.org/10.6035/Sapientia93>
- Mendenhall, W., Beaver, R., y Beaver, B. (2023). *Introducción a la probabilidad y estadística* <https://latam.cengage.com/libros/introduccion-a-la-probabilidad-y-estadistica-de-mendenhall-2023/>
- Salazar, C. y Del Castillo, S. (2018). *FUNDAMENTOS BÁSICOS DE ESTADÍSTICA*. <http://librodigital.sangregorio.edu.ec/librosusgp/B0009.pdf>

ANEXOS

Anexo 1:

https://colab.research.google.com/drive/1gEUwi81sezPRM2dr_77T7fAunUL1W1eI?usp=sharing

Anexo 2:

https://colab.research.google.com/drive/1CRmid6r5klqo3Y_OvdNGFI670CPXHk0Q?usp=sharing

Plataforma Datos abiertos Ecuador:

<https://datosabiertos.gob.ec/>

SRI_Vehiculos_Nuevos_2022:

<https://www.datosabiertos.gob.ec/dataset/estadisticas-vehiculos-2022>

inocar_altura_horaria_predicha_2023diciembre:

<https://www.datosabiertos.gob.ec/dataset/altura-horaria-predicha-de-mareas>

Drive con las bases de datos utilizadas:

<https://drive.google.com/drive/folders/1ajIODY8adSatRBsaE1-wtRRftgd3q9XY?usp=sharing>

Anexo 2. Oficio de Apertura de la Institución y certificado de la ejecución de experiencia áulica.



UNL

Universidad
Nacional
de Loja

Carrera de Pedagogía de las
Ciencias Experimentales:
Matemáticas y la Física

Loja, 05 de abril del 2024

Magister.

Juan Luna Rengel.

RECTOR UNIDAD EDUCATIVA "TCRN. LAURO GUERRERO"

Ciudad. -

Mediante la presente, la Carrera de Pedagogía de las Ciencias Experimentales: Matemáticas y Física de la Facultad de la Educación, el Arte y la Comunicación de la Universidad Nacional de Loja, le extiende un cordial saludo y nos permitimos solicitarle lo siguiente:

Acudimos a su autoridad, para solicitarle de la manera más comedida se le autorice a quien corresponda las facilidades para que nuestro estudiante **Jean Carlo Prado Vélez, del Ciclo VIII**, pueda realizar un acercamiento a la institución en la cual se llevará a cabo una experiencia áulica en el que se aplicara el uso del software Python, dado que esto es necesario para que dicho estudiante pueda realizar el Trabajo de Integración Curricular (Tesis), con el tema "Uso de datos abiertos para la enseñanza de la estadística descriptiva en el nivel secundario con Python".

Esperando su valiosa contribución a la formación de profesionales, me suscribo de usted no sin antes expresarle mis sentimientos de consideración y estima personales.

Atentamente,

AUTORIZADO, por favor coordinar con el Vicerrector y el área de matemática.

Loja, 26-04-2024

PhD. Ángel Klever Orellana Malla.

DIRECTOR DE LA CARRERA DE PEDAGOGÍA
DE LAS CIENCIAS EXPERIMENTALES: MATEMÁTICAS Y LA FÍSICA



c.c. archivo de la carrera
Elaboración Lcdo. Alberto Miguel Carrión.

Educamos para Transformar



unl

Universidad
Nacional
de Loja

Facultad
de la Educación,
el Arte y la Comunicación

Magister.
Juan Luna Rengel.

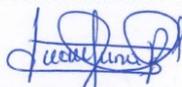
RECTOR DE LA UNIDAD EDUCATIVA "TCRN. LAURO GUERRERO"

CERTIFICO:

Que el señor, **Jean Carlo Prado Vélez** con cédula **1105484404**, estudiante de VIII Ciclo de la carrera de Pedagogía de las Ciencias Experimentales: Matemáticas y la Física, Facultad de la Educación, el Arte y la Comunicación, de la Universidad Nacional de Loja, realizó un total de 6 clases demostrativas en el marco del Trabajo de Integración Curricular: **Uso de datos abiertos para la enseñanza de la Estadística Descriptiva en el nivel secundario con Python**, las cuales se desarrollaron dentro de la institución desde el 20/05/2024 hasta el 30/05/2024.

Durante el periodo señalado el estudiante ha cumplido satisfactoriamente con las tareas y exigencias acordadas en el Plan de contenidos, por lo tanto, se certifica que ha culminado satisfactoriamente las actividades previstas.

Loja, 03 de Junio de 2024


Mg. Juan Luna Rengel.
RECTOR



Ciudadela Universitaria "Pío Jaramillo Alvarado",
Sector La Argelia · Loja - Ecuador
072-54 7234

Anexo 3. Bitácora de búsqueda

| Criterios de búsqueda: | Criterios de selección: |
|---|--|
| Palabras clave: “datos abiertos”, “enseñanza de estadística descriptiva”, “nivel secundario”, “Python”, “educación”, “recursos didácticos”, etc | Relevancia del tema: el documento aporta información útil |
| Fecha de publicación: Documentos publicados en los últimos 5 años, últimos 10 años, últimos 20 años | Calidad del documento: El documento está bien escrito, el documento presenta argumentos sólidos, el documento es de una fuente confiable |
| Tipo de documento: Artículos de revistas académicas, tesis de maestría, tesis de doctorado, libros | Autoridad del autor: El autor tiene experiencia en el campo, el autor tiene credenciales académicas, el autor ha publicado otros trabajos sobre el tema |
| Idioma del documento: Documentos en español, documentos en inglés, documentos traducidos al español | Accesibilidad del documento: El documento está disponible en línea, el documento está disponible en la biblioteca, el documento está disponible en una base de datos académica, el documento está disponible en formato PDF. |
| Lugar de publicación: Documentos publicados en el país o documentos publicados en países de habla hispana, | Claridad de la presentación: El documento está bien organizado, el documento presenta las ideas claramente, el documento utiliza un lenguaje comprensible. |
| Autores: Documentos escritos por expertos en el campo (Carmen Batanero) | Posición en los resultados de búsqueda: El documento apareció entre los primeros 10 a 15 resultados de la búsqueda, lo que puede indicar su relevancia y popularidad en el tema. |
| Formato: Documentos en PDF, documentos en HTML, documentos en formato de texto. | |
| Disponibilidad: Documentos disponibles gratuitamente, documentos disponibles en bibliotecas | |

| Motor de búsqueda | Ecuación de búsqueda | N. documentos | N. documentos seleccionados | Tipo de documento | Año | Autor/es | Título y doi o url |
|-------------------|--|---------------|-----------------------------|-------------------|------|---|--|
| Scopus | Python AND statistics AND secondary AND education | 12 | 4 | Artículo | 2022 | Yannik Fleischer, Rolf Biehler y Carsten Schulte | Teaching and learning data driven machine learning with educationally designed jupyter notebooks https://iase-web.org/ojs/SERJ/article/view/61 |
| | | | | Artículo | 2020 | Marc Tang | A simple introduction to programming and statistics with decision trees in R https://doi.org/10.1111/test.12210 |
| | | | | Artículo | 2022 | Susanne Podworny, Sven Hüsing, Carsten Schulte | A place for a data science introduction in school: between statistics and programming https://iase-web.org/ojs/SERJ/article/view/46 |
| | | | | Artículo | 2016 | Ching Chang, Ya-Lun Chin, Chih-Kai Chang | Experimental Functionality Development for Scratch Mathematical and Statistics Extensions https://ieeexplore.ieee.org/document/7858553 |
| Scopus | (TITLE-ABS-KEY (teaching of statistics) AND TITLE-ABS-KEY (with Python ion secondary education)) | 8 | 2 | Artículo | 2022 | Fabricio Vines y Wilmer Ríos | Cuadernos interactivos: experiencia de enseñanza y aprendizaje de estadística en el entorno Jupyter |
| | | | | Artículo | 2022 | Schanzer, E., Pfenning, N., Denny, F., Dooman, S., Politz, J., Lerner, B., Fisler, K., y Krishnamurthi, S | Integrated Data Science for Secondary Schools: Design and Assessment of a Curriculum https://doi.org/10.1145/3478431.3499311 |
| Google Académico | Teaching of statistics AND Python AND Data Science AND secondary education | 29 | 5 | Artículo | 2021 | Marlena Duda, Kelly L. Sovacool, Negar Farzaneh1, Vy Kim Nguyen1, Sarah E. Haynes, Hayley Falk1, Katherine L. Furman3, Logan A. Walker, Rucheng Diao, Morgan Oneka, Audrey C. Drotos, Alana Woloshin, Gabrielle A. Dotson, April Kriebel, Lucy Meng, Stephanie N. | Teaching Python for Data Science: Collaborative development of a modular & interactive curriculum https://doi.org/10.21105/jose.00138 |

| | | | |
|----------|------|---|--|
| | | Thiede, Zena Lapp, and Brooke N. Wolford | |
| Articulo | 2020 | Biehler, R., Fleischer, Y., Budde, L., Frischemeier, D., Gerstenberger, D., Podworny, S., y Schulte, C. | Data science education in secondary schools: Teaching and learning decision trees with CODAP and jupyter notebooks as an example of integrating machine learning into statistics education https://doi.org/10.52041/SRAP.20304 |
| Articulo | 2019 | Bryson Stemock y Lucy Kerns | Use of commercial and free software for teaching statistics. https://doi.org/10.52041/serj.v18i2.140 |
| Articulo | 2023 | Justin O. Holman y Allie Hacherlb | Teaching Monte Carlo Simulation with Python. . https://doi.org/10.1080/26939169.2022.2111008 |
| Articulo | 2023 | Jana Lasser, Debsankha Manik, Alexander Silbersdorff y Benjamin Säfken | Introductory data science across disciplines, using Python, case studies, and industry consulting projects https://onlinelibrary.wiley.com/doi/full/10.1111/test.12243 |

Anexo 4. Certificado de traducción del resumen



Loja, 30 de julio de 2024

Leda, Ana Lucia Contento Japón
LICENCIADA EN PEDAGOGÍA DEL IDIOMA INGLÉS

CERTIFICO:

Que el resumen del Trabajo de Integración Curricular cuyo título es: **Uso de datos abiertos para la enseñanza de la estadística descriptiva en el nivel secundario con Python**, del aspirante **Jean Carlo Prado Vélez**, con cédula de identidad Nro. **1105484404** ha sido traducido al inglés y cumple con las características propias del idioma extranjero.

Resumen:

En el siglo XXI, la Estadística es importante para la toma de decisiones informadas y su enseñanza en la educación es esencial, especialmente cuando se emplean herramientas digitales como Python para manejar datos reales. Por ello, esta investigación busca analizar el proceso de enseñanza de la Estadística Descriptiva en nivel secundario utilizando datos abiertos gestionados con Python, para esto se realizó un proceso riguroso de revisión documental sistemática, en artículos, y capítulos de libros. La búsqueda de información se efectuó en bases de datos científicas y motores de búsqueda, las técnicas para selección y organización se realizaron mediante bitácoras de búsqueda, fichas bibliográficas y de contenido. Con base en esa teoría se elaboró 6 talleres abordados como clases en una institución Fiscomisional de la ciudad de Loja, donde se abordó temáticas de Estadística Descriptiva haciendo uso de datos abiertos y Python, siguiendo una secuencia de Inicio (3 talleres), Desarrollo (2 talleres) y Cierre (1 taller), finalmente se aplicó un instrumento para medir el nivel de satisfacción de los estudiantes. Como resultados se obtuvo un proceso de enseñanza adecuado para esta disciplina y un grado de satisfacción alto. Gracias a esto se concluyó que los estudiantes evidencian quedar satisfechos, lo que implica que el uso de la secuencia de las 3 fases y la inclusión de una herramienta digital genera cierta predisposición en el alumno para aprender Estadística Descriptiva.

Palabras clave: Herramientas digitales, Software estadístico, Estadística aplicada

Educamos para **Transformar**



UNL

Universidad
Nacional
de Loja

Abstract:

In the 21st century, Statistics is important for informed decision making and its teaching in education is essential, especially when digital tools such as Python are used to handle real data. Therefore, this research seeks to analyze the teaching process of Descriptive Statistics at the secondary level using open data managed with Python, for this a rigorous process of systematic documentary review was carried out, in articles, and book chapters. The search for information was carried out in scientific databases and search engines, the techniques for selection and organization were carried out by means of search logs, bibliographic and content files. Based on this theory, 6 workshops were developed as classes in a Fiscomisional institution in the city of Loja, where Descriptive Statistics topics were addressed using open data and Python, following a sequence of beginning (3 workshops), Development (2 workshops) and Closing (1 workshop), finally an instrument was applied to measure the level of satisfaction of the students. The results showed an adequate teaching process for this discipline and a high degree of satisfaction. Thanks to this, it was concluded that the students were satisfied, which implies that the use of the sequence of the 3 phases and the inclusion in the student to learn Descriptive Statistics

Keywords: Digital tools, Statistical software, Applied statistics

Lo certifico en honor a la verdad.

Lcda. Ana Lucia Contento Japón

LICENCIADA EN PEDAGOGÍA DEL IDIOMA INGLÉS

Educamos para Transformar

