|  |  |
| --- | --- |
| PROGRAMA DE FORMACIÓN | Gestión de analítica de datos |

|  |  |  |  |
| --- | --- | --- | --- |
| COMPETENCIA | 210601026: Procesar datos de acuerdo con el procedimiento técnico y la metodología estadística. | RESULTADOS DE APRENDIZAJE | 210601026-01. Recopilar datos de acuerdo con la metodología estadística. |

|  |  |
| --- | --- |
| NÚMERO DEL COMPONENTE FORMATIVO | CF2 |
| NOMBRE DEL COMPONENTE FORMATIVO | Estadística básica |
| BREVE DESCRIPCIÓN | Este componente formativo, aborda generalidades y aspectos claves de la estadística básica. Se utilizan conceptos fundamentales como: población, muestra, submuestra, censo, tipos de muestreo, variables cualitativas, variables cuantitativas. Con el estudio de este componente, el aprendiz se afianzará en todo lo relacionado con datos, análisis estadístico, diseño de instrumentos, recolección de datos y verificación de la calidad. |
| PALABRAS CLAVE | Estadística básica, población, muestra, tipos de muestreo, tipos de variables. |

|  |  |
| --- | --- |
| ÁREA OCUPACIONAL | Ciencias sociales, educación, servicios gubernamentales y religión. |
| IDIOMA | Español |

**Tabla de contenido**

1. [**Estadística básica**](#_heading=h.26in1rg)
2. [**Tipos de variables**](#_heading=h.lnxbz9)
3. [**Escalas de medición**](#_heading=h.35nkun2)
4. [**Diseño de instrumentos, recolección de datos, verificación de la calidad**](#_heading=h.2jxsxqh)

# Introducción

|  |
| --- |
| Cuadro de texto |
| Estimado aprendiz, bienvenido a este componente formativo, donde se profundizan los elementos más importantes para la comprensión, asimilación y aplicación de la estadística básica, en la dinámica de las empresas y organizaciones, en general.  Conviértase en experto y aplique los conceptos fundamentales que configuran la temática, identifique los elementos constitutivos de la misma, aprópiese del concepto y de las acciones de medición, recolección de información, diseño de instrumentos para recolectarla, verificación de la calidad, entre otros.  En el siguiente video conocerá, de forma general, la temática que se estudiará a lo largo del componente formativo. |

**Guion de video introductorio**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Tipo de recurso** | Video spot animado | | | |
| **NOTA** | **La totalidad del texto locutado para el video no debe superar las 500 palabras aproximadamente** | | | |
| **Título** |  | | | |
| **Escena** | **Imagen** | **Sonido** | **Narración (voz en off)** | **Texto** |
| **1** | 228131\_i201 | NA | En este momento con el auge de las TIC y la digitalización, l a sociedad genera datos diariamente de forma intensiva. | La digitalización. La sociedad genera datos diariamente de forma intensiva. |
| **2** | Banca en línea en teléfonos inteligentes  228131\_i202 | NA | Esto se hace mediante transferencias electrónicas, compras en línea, interacciones en las redes sociales, entre otros. | Transferencias electrónicas, compras en línea, interacciones en las redes sociales. |
| **3** | Ilustración del concepto de análisis  228131\_i203 | NA | Lo que constituye el origen para los diferentes análisis en las empresas. | Análisis de datos en las empresas. |
| **4** | Ilustración de concepto informe de datos  228131\_i204 | NA | Las compañías, hoy en día, se han enfocado en esa salida enorme de datos producidos en el mundo | Empresas enfocadas en la gran cantidad de datos. |
| **5** | Lupa con búsqueda de archivos  228131\_i205 | NA | con la finalidad de obtener esa información cruda | Obtener la información cruda |
| **6** | Plantilla de infografía  228131\_i206 | NA | para ordenarla, limpiarla, procesarla y analizarla para la futura toma de decisiones en las empresas, | Limpiarla, procesarla y analizarla para tomar decisiones. |
| **7** | Ilustración de concepto carga rápida  228131\_i207 | NA | b uscando optimizar su producto, teniendo siempre presente que esto puede distinguirlos de la competencia, mejorando sus ganancias. | Optimizar su producto |
| **8** | Hombre con laptop analizando infografías. diagrama, gráfico de barras, informe ilustración vectorial plana. análisis, marketing, director de proyectos  228131\_i208 | NA | Para ello, buscan personal que se incluya en este universo de la información | Personal que se sumerge en este universo de la información. |
| **9** | Sí o no ilustración del concepto  228131\_i209 | NA | con el objetivo final de tener una fundamentación teórica para aportar en la futura toma de decisiones, | Toma de decisiones |
| **10** | Plantilla de panel de usuario de interfaz de tableroI  228131\_i210 | NA | utilizando herramientas analíticas y su visualización mediante tableros de mando *(dashboard)* | Herramientas analíticas y su visualización mediante tableros de mando *(dashboard).* |
| **11** | Generación de ideas de negocio. desarrollo del plan. hombre pensativo con personaje de dibujos animados de bombilla. mentalidad técnica, mente emprendedora, proceso de lluvia de ideas.  228131\_i211 | NA | o contando historias mediante la narración gráfica *(storytelling),* del problema a resolver para la empresa. | Narración gráfica *(storytelling)* del problema a resolver. |
| **12** | Big data center, rack de sala de servidores, proceso de ingeniería, trabajo en equipo, tecnología informática, almacenamiento en la nube  228131\_i212 | NA | Es tal la cantidad de datos que, actualmente, han salido nuevos cargos laborales alrededor de los mismos . | Nuevos cargos laborales alrededor de los datos. |
| **13** | Analista que trabaja con Business Analytics y Data Management System en el equipo para realizar informes con KPI y métricas conectadas a la base de datos. Estrategia corporativa para finanzas, operaciones, ventas, marketing - Foto de stock de Datos libre de derechos  228131\_i213 | NA | Científico de datos, ingeniero de datos, *machine learning* ingeniero, arquitecto de datos, *data mánager*, *chief data officer*, analista de datos, entre otros. | Científico de datos, ingeniero de datos, *machine learning* ingeniero, arquitecto de datos. |
| **14** | Tecnología de base de datos y red. - Foto de stock de Base de datos libre de derechos  228131\_i214 | NA | Las industrias deben tomar esta gran cantidad de información y organizarla en una base de datos. | Organizar y almacenar en bases de datos. |
| **15** | Concepto de alojamiento web con personajes de personas. Base de datos en línea, servidor, centro de datos web, computación en la nube, tecnología, computadora, seguridad. Ilustración vectorial moderna aislada para web, banner, póster, ui - arte vectorial de Maestro de ceremonias libre de derechos  228131\_i215 | NA | Mediante el proceso de extraer, transformar y cargar la información en una base de datos que contendrá esta información procesada. | Extraer, transformar y cargar los datos. |
| **16** | Business Analytics and Data Management System (DMS) proporcionando información clave para la estrategia corporativa. Concepto con analista experto construyendo visualización con KPI y métricas a partir de la información de la base de datos - Foto de stock de Base de datos libre de derechos  228131\_i216 | NA | A partir de tal información se hacen los análisis correspondientes. | Hacer los análisis |
| **17** | encontrar la solución para un concepto de problema con piezas de rompecabezas - Foto de stock de Solución libre de derechos  228131\_i217 | NA | Además, resuelven problemas que le interesan a las empresas con diferentes ámbitos como fuga de cliente. | Resolver problemas que le interesan a las empresas. |
| **18** | gráfico de inversión en el mercado bursátil con datos de indicador y volumen. - Foto de stock de Datos libre de derechos  228131\_i218 | NA | El propósito de este componente específico es ubicar y ayudar al aprendiz a distinguir los conceptos y términos más relevantes en la estadística básica. | Ayudar al aprendiz a distinguir los conceptos y términos más relevantes en la estadística básica. |
| **19** | Vista aérea de la multitud conectada por líneas - Foto de stock de Conexión libre de derechos  228131\_i219 | NA | Términos como escala de medición, cuestionarios,  encuestas, verificación de la  calidad. | Población, muestra, submuestra, estadístico, tipos de muestreo. |
| **20** | Lista de verificación de papel y pictograma plano lápiz - arte vectorial de Lista de chequeo libre de derechos  228131\_i220 | NA | a sí como, tipos de variables, cualitativas, cuantitativas, diseño de instrumentos, recolección de datos entre otros. | Diseño de instrumentos, recolección de datos. |
| **21** | Estadísticas de diseño de vectores planos y análisis de datos para el concepto de inversión en finanzas empresariales con el equipo de empresarios que trabaja en el panel de control de gráficos de monitor - arte vectorial de Datos libre de derechos  228131\_i221 | NA | Todo lo anterior, para la fundamentación teórica en analítica de datos | Analítica de datos |
| **22** | Granjero de sexo femenino feliz trabajando en la granja para alimentar a la población ilustración vectorial plana. granja de dibujos animados con tecnología de automatización.  228131\_i222 | NA | y su aplicación de manera efectiva en la solución de problemas reales. | Solución de problemas reales. |
| **23** | Dardo de pantalla de impresión y tablero de destino en las flechas hacia arriba. Objetivo de inversión y concepto de negocio. - Foto de stock de Aspiraciones libre de derechos  228131\_i223 | NA | El componente entonces se desarrolla de forma gradual iniciando por las bases estadísticas; | Bases estadísticas |
| **24** | Mujer con portátil, estudio o concepto de trabajo. Mesa con libros, lámpara, taza de café. Ilustración vectorial en estilo plano - arte vectorial de Aprender libre de derechos  228131\_i224 | NA | cuya metodología es interactiva, utilizando diferentes herramientas audiovisuales con sus respectivas explicaciones teóricas. | Interactiva utilizando diferentes herramientas audiovisuales. |
| **25** | Chica escribiendo en diario o diario - arte vectorial de Lápiz libre de derechos  228131\_i225 | NA | Aquí es muy importante el trabajo individual de cada aprendiz para maximizar la curva de aprendizaje. | Maximizar la curva de aprendizaje. |
| **26** | Equipo de Hombres Programadores que trabajan en Desarrollo Web.Brainstorming Process.Script Coding,Programación en php,python, javascript Lenguajes artificiales. Desarrollador de software. Ilustración de dibujos animados vectoriales planos. - arte vectorial de Programador de informática libre de derechos  228131\_i226 | NA | En este componente se trabaja con *Google Colab*, el cual permite ejecutar código en el lenguaje *Python*, sin tener que instalarlo. | Lenguaje Python |
| **27** | Descargar CSV en la pantalla del ordenador portátil. Concepto de documento para descargar. Archivos con etiqueta CSV y abajo el signo de flecha. Ilustración de vector. - arte vectorial de Descargar - Internet libre de derechos  228131\_i227 | NA | Se utilizarán datos en formato CSV, tomados de unos créditos de consumo otorgados en Medellín, | Datos en formato CSV |
| **28** | Grandes datos - arte vectorial de Fichero archivador libre de derechos  228131\_i228 | NA | o btenidos de la página de datos abiertos de la Alcaldía de Medellín, MEData. | Datos abiertos |
| **29** | Un gráfico circular con personas.  228131\_i229 | NA | En este componente también aprenderá a distinguir conceptos fundamentales como: población, muestra, submuestra, censo, estadísticos. | Población, muestra, submuestra, censo, estadísticos. |
| **30** | Iconos de infografía de población de flechas  228131\_i230 | NA | Así como, tipos de muestreo,con sus respectivas variables cualitativas y cuantitativas. | Variables cualitativas, variables cuantitativas. |
| **31** | Ilustración del concepto de encuesta al cliente  228131\_i231 | NA | Igualmente, el diseño de instrumentos resolviendo algún problema en particular. | Instrumento, recolección de datos. |
| **32** | Business people team analytics Datos e investigación en el panel de informes de monitoreo web supervisan el concepto y el gráfico de inversión financiera reportan los resultados de rendimiento con el concepto de trabajo de reunión de negocios - arte vectorial de Big Data libre de derechos  228131\_i232 |  | Con la finalidad de tener la fundamentación teórica para la compresión del análisis estadístico | Compresión del análisis estadístico. |
| **33** | Ilustración del concepto de calidad del producto  228131\_i233 |  | Utilizando todas estas bases para un estudio de calidad y fundamentado en la teoría para llevarlo a la práctica. | Estudio de calidad |
| **34** | 228131\_i234 | NA | Se reforzarán términos relacionados con estadística y su aplicación en encuestas como instrumento de recolección. | Reforzará en términos como estadística y su aplicación en encuestas como instrumento de recolección |
| **Nombre del archivo** | **228131\_v01** | | | |

**Desarrollo de contenidos**

|  |
| --- |
| Cuadro de texto |
| ~~Se invita a desarrollar todo el contenido, al igual que a realizar las actividades de aprendizaje; de esa forma, al finalizar el estudio de desarrollo temático, se habrán alcanzado las competencias en estadística básica y se estará en capacidad de recopilar datos de acuerdo con la metodología estadística.~~ |

# Estadística básica

|  |  |  |
| --- | --- | --- |
| **Tipo de recurso** | | Pestañas o tabs Verticales |
| **Introducción** | | Para adentrarse en el concepto y aplicación de la estadística básica, es importante y fundamental, conocer y comprender algunos aspectos conceptuales como: población, muestra, submuestra censo, parámetros, estadísticos, tipos de muestreo: aleatorio simple, estratificado, sistemático.  Comprenda la definición inicial de cada uno de ellos, a continuación, y procure tomar nota de lo más destacado: |
| **Imagen:** 228131\_i235 | | |
| **Población** | **Población**  Una población en estadística está relacionada con el número de individuos o elementos que se van a estudiar o analizar, estos individuos pueden ser personas y características**.**  **Ejemplos:**  La población de estudiantes de la universidad colombiana.  La población de los beneficiarios del subsidio para el adulto mayor.  La población de estudiantes del Sena en Colombia. | |
| **Muestra** | **Muestra y submuestra**  Una muestra se obtiene de una población cuando se saca un subconjunto de esa población. Por ejemplo: de la población de estudiantes Sena, los que pertenecen al Sena Regional Antioquia.  Una submuestra es tomar una muestra a la muestra. Por ejemplo: de la muestra Sena Regional Antioquia tomar el Sena Antioquia del Municipio de Bello. | |
| **Censo** | **El censo**  En estadística es el conteo de individuos que conforman una población**.** Por ejemplo:el censo del DANE, que consiste en ir casa por casa, para contar y caracterizar a los individuos que viven en Colombia. | |
| **Estadístico** | **Estadísticos**  Un estadístico es un valor que se calcula a partir de los datos de una muestra. La finalidad es representar una característica de la muestra e inferir la misma característica sobre la población.  **Ejemplo**: de la población de estudiantes del Sena se obtiene una muestra de la regional Valle y se calcula el promedio de edad de los estudiantes regional Valle. Los cuartiles (Q1, Q2, Q3) de la edad de los estudiantes regional Valle. | |
| **Parámetro** | **Parámetro**  Es un valor numérico constante, el cual describe una característica de la población.  **Ejemplo**: el promedio de edad de estudiantes del Sena, la desviación del peso de los estudiantes del Sena Regional Antioquia, la moda del sueldo mensual de los docentes del Sena Regional Cundinamarca.  Es importante tener en cuenta que el parámetro describe una característica de la población, mientras que el estadístico describe una característica de la muestra. Además, el estadístico se utiliza para aproximar un parámetro. | |

**Tipos de muestreo**

|  |  |
| --- | --- |
| **Tipo de recurso** | Acordeón tipo 1 |
| **Introducción** | El muestreo es un proceso o conjunto de métodos, para obtener una muestra finita de una población finita o infinita para estimar valores.  A continuación, se pueden conocer los tipos de muestreo: |
| **Imagen:** 228131\_i236 | |
| **Muestreo aleatorio simple**  Es una técnica de muestreo donde todos los individuos que conforman la muestra tienen la misma probabilidad de ser elegidos. Si el individuo puede ser seleccionado más de una vez estamos en presencia de un muestreo aleatorio simple con reposición, en caso contrario, sería muestreo aleatorio simple sin reposición.  **Ejemplo:**  Se tiene una población de 500 personas, se quiere sacar una muestra de 50 personas. Se asigna a las 500 personas un número y, de la misma forma que un sorteo de lotería, selecciona 50 números al azar. Es así como, la totalidad de estos números al azar sería la muestra de 50 personas. | |
| **Muestreo aleatorio estratificado**  Este tipo de muestreo se utiliza cuando se quiere saber la relación entre los subgrupos o estratos por alguna característica o múltiples características en común. Por ejemplo: el año cuando fue otorgado un crédito, la edad de la persona que recibe el crédito, etc.  **Ejemplo:** se puede dividir la muestra de las personas a quienes les fue otorgado un crédito de consumo, en 7 subgrupos por edad, cada letra corresponde a un grupo: a) 18-27 años, b) 28-37 años, c) 38-47 años, d) 48-57 años, e) 58-67 años, f) 68-77 años, g) 78-88 años. | |
| **Muestreo sistemático**  Se utiliza cuando el muestreo aleatorio simple no se puede llevar a cabo por algún motivo.  **Ejemplo:**  De la población que se tiene 12152 se quiere una muestra de 500 elementos, para ello se divide el total de la población en 50 pedazos o fragmentos aleatorios de 10 individuos. Luego se selecciona un número al azar entre 1 y 10.  Elegir de forma aleatoria el primer individuo, después definir el tamaño de un intervalo para seleccionar los otros individuos que hacen parte de la muestra. Este tamaño del intervalo lo define la persona a cargo de la investigación.  **Ejemplo:**  Se va a extraer una muestra de 10 personas de una muestra de 50 personas. El primer individuo seleccionado para la muestra es el número 5 con un tamaño o intervalo de 4, decidido por el investigador.  Los elementos de la muestra serían: 5, 9, 13, 17, 21, 25, 29, 33, 37, 41. | |

|  |
| --- |
| Cuadro de texto |
| Para reforzar los conceptos sobre este tema, se invita a leer el libro digital Estadísticade MarioTriolla de 2018, que se encuentra en el material complementario y puede ser consultado en la biblioteca digital SENA. |

# Tipos de variables

|  |
| --- |
| Cuadro de texto |
| Son múltiples los tipos de variables que se trabajan en los procesos estadísticos: cualitativas o categóricas, cuantitativas o numéricas, discretas y continuas. |

|  |  |
| --- | --- |
| **Tipo de recurso** | Rutas / Pasos. Verticales 1 |
| **Introducción** | A continuación, se explican los diferentes tipos de variables con un ejemplo real, con gráficos generados dentro de Google Colab, utilizando el lenguaje Python: |
| **Imagen:** 228131\_i237 | |
| **Botón 1** | **Obtener los datos**  Para estos gráficos, se tomaron unos datos de créditos de consumo, otorgados en Medellín, obtenidos de la página de datos abiertos de la Alcaldía de Medellín, en formato CSV. |
| **Botón 2** | **Importar librería**  Se realiza dentro de Google Colab, importando la librería panda, mediante **import pandas as pd**, que es la abreviación de pandas. |
| **Botón 3** | **Lectura del archivo CSV**  Para leer el archivo se utiliza pd.read\_csv(‘nombre del archivo’, sep=’;’). En este caso, el nombre del archivo es créditos\_de\_consumo\_otorgados.csv’ y está separado por un signo de punto y coma (**;).**  Un archivo CSV (valores separados por comas). Es un archivo de texto que tiene un formato específico, que permite guardar los datos en un formato de tabla estructurada.  Un archivo CSV puede tener diferentes separadores, distintos a la coma (,) como:  Punto y coma (;), barra oblicua ( / ), guión del medio (-), guion de piso (\_), etc. |
| **Botón 4** | **Obtener las primeras cinco filas**  La sentencia df.head (5) nos permite ver las primeras 5 filas del **data frame***,* en este caso se llamó **df** puede tener cualquier nombre. |

|  |  |  |
| --- | --- | --- |
| **Tipo de recurso** | Slider Presentación | |
| **Introducción** | Un data frame es una estructura de datos con dos dimensiones (filas y columnas), se pueden guardar datos de distintos tipos (caracteres, enteros, valores de punto flotante, etc.) en columnas. Es similar a una hoja de cálculo o una tabla de SQL.  Continuando con el ejemplo de créditos otorgados, se deben observar, a continuación, los gráficos que ilustran las operaciones realizadas sobre la data: | |
| El data frame que se observa en la imagen, contiene las primeras filas de la data frame*,* con un total de 12 columnas.  Las columnas son:  ['edad', 'sexo', 'cabeza\_de\_hogar', 'caracterizacion\_ciudadano', 'estrato', 'barrio\_vereda', 'comuna', 'zona', 'fecha\_de\_beneficio\_dia','fecha\_de\_beneficio\_mes', 'fecha\_de\_beneficio\_año', 'monto\_prestado'.  **¡Importante!**  Estos nombres los asignó la persona que subió el archivo en formato CSV a **Metadata** alcaldía de Medellín, 2022. Generalmente hay datos faltantes, los nombres y valores de las columnas tienen errores de digitación, ortográficos, etc. Por esta razón, **siempre se debe hacer una limpieza y depuración de los datos.** | | **~~Figura 1~~**Créditos de consumo.    **Nota**.Muestra las columnas de los créditos de consumo otorgados.  **Imagen:** 228131\_i238 |
| Los datos faltantes tienen diferentes tratamientos dependiendo de: los datos, la cantidad de valores faltantes y el tipo de problema a resolver. Se debe utilizar un criterio para tratarlos. Por ejemplo, sustituirlos por el promedio de la columna, sustituirlo por el valor de la columna que está por encima o por debajo (dependiendo de la cantidad de datos faltantes se debe analizar con detalle), sustituirlo por el valor que más se repite (moda), sustituirlo por la mediana; si son pocos se podría eliminar previo a un análisis riguroso. | | Ubicar la misma figura anterior |
| Las columnas del archivo créditos de consumo, tienen diferentes tipos de datos. | | **~~Figura 2~~**Tipos de datos.    **Nota.** Muestra información sobre los tipos de datos de las columnas.  **Imagen:** 228131\_i239 |

|  |
| --- |
| Cuadro de texto |
| ~~Hasta ahora se han estudiado los conceptos básicos de estadística y los tipos de datos. A continuación, se abordan las escalas de medición.~~ |

# Escalas de medición

|  |  |  |
| --- | --- | --- |
| **Tipo de recurso** | Slider Presentación | |
| **Introducción** | Uno de los elementos básicos en la definición de una variable es el tipo de escala utilizada para medirla. En función de la escala seleccionada, se decide su codificación, tratamiento informático y estadístico.  Hay varios tipos de escalas de medición que, a continuación, se desarrollan y ejemplifican: | |
| **Categóricos**  *object y booleano* (booleanos o lógicos, verdadero o falso, *true, false*).  **Numéricos**  *int64*, significa valores enteros o discretos, representa un entero de 64 bits.  **Numéricos**  *float64* (continuos), significa que toman valores reales de 64 bits.  En 64 bits, el rango es de -9.223.372.036.854.775.808 hasta 9.223.372.036.854.775.807. | | Colocar imagen relacionada con el texto    **Imagen:** 228131\_i240 |
| **Edad es una variable numérica**  En tanto variable numérica, puede ser entera o discreta, tomando, por ejemplo, valores como:70, 55, 53, 45, 28, etc.  *dtype:int* 64**,** significa el tipo de dato, en este caso toma valores enteros de 64 bits.  En la figura se observa la edad, con su frecuencia en orden descendente; es decir, desde la mayor frecuencia en edad y, en este caso, para 70 años tiene una frecuencia de 514 individuos o datos. Le sigue una edad de 55 años con una frecuencia de 290 individuos o datos, y así sucesivamente. | | **~~Figura 3~~**  Frecuencia de edades en orden descendente.    ***~~Nota.~~*** *~~Muestra la edad con su frecuencia en orden descendente.~~*  **Imagen:** 228131\_i241 |
| **Categórico ordinal**  Valores que representan categorías con alguna clasificación intrínseca, tienen características de orden. Por ejemplo: malo, bueno, excelente.  **Categórico nominal**  Representan categorías sin clasificación intrínseca. Por ejemplo, las marcas de zapatos, género, raza, etnia.  La edad (en años) puede ser tratada así:   * Nivel ordinal: (niño/adolescente/joven). * Nominal: (adulto/no adulto). | | **Imagen:** 228131\_i242 |
| **Caracterización del ciudadano**  Es un dato categórico nominal, toma valores como: otro, independiente, emprendedor, empresario, etc**.**  En la figura se observa la caracterización del ciudadano con su frecuencia en orden descendente. Es decir, desde la mayor frecuencia la caracterización del ciudadano y, en este caso, para otros, tiene una frecuencia de 3525 individuos o datos, le sigue independiente con una frecuencia de 2885 individuos o datos, y así sucesivamente. | | **~~Figura 4~~**  Caracterización del ciudadano con frecuencia en orden descendente.    ***~~Nota.~~*** *~~Muestra los valores de la columna Caracterización del ciudadano, con su frecuencia en orden descendente.~~*  **Imagen:** 228131\_i243 |
| **Cabeza de hogar**  Es un dato categórico nominal, toma los valores: no, sí, aporta económicamente, 0, solidaria.  En la figura se observa la columna cabeza de hogar con su frecuencia en orden descendente. Es decir, desde la mayor frecuencia cabeza de hogar y, en este caso, para el valor “no” tiene una frecuencia de 6413 individuos o datos, le sigue el valor “sí” con una frecuencia de 3987 individuos o datos, y así sucesivamente. | | **~~Figura 5~~**  Frecuencias cabeza de hogar.    **~~Nota.~~** *~~Muestra los valores de la columna Cabeza de hogar con su frecuencia en orden descendente.~~*  **Imagen:** 228131\_i244 |
| **La columna sexo**  Correspondiente al género, es una variable categórica nominal, toma los valores: femenino y masculino.  En la figura se observa que, para el valor femenino, se tiene una frecuencia de 8069, casi el doble, del valor masculino, que tiene una frecuencia de 4083. | | **~~Figura 6~~**  Valores de columnas nombre y sexo, correspondiente al género.    ***~~Nota.~~*** *~~Muestra los valores de la columna de nombre sexo correspondiente al género.~~*  **Imagen:** 228131\_i245 |
| **La fecha del beneficio**  La fecha de beneficio del mes es un valor numérico entero o discreto, toma valores 1, 2, 3,12 correspondientes al número del mes: 1 es enero, 2 es febrero y así sucesivamente**.**  En la figura se observa la columna, fecha de beneficio mes, con su frecuencia en orden descendente. Es decir, desde la mayor frecuencia en la columna fecha de beneficio mes; en este caso, para el mes 5 (mayo), tiene una frecuencia de 2490 individuos o datos y les fueron otorgados el crédito de consumo; le sigue el valor para el mes 6 (junio) con una frecuencia de 1950 individuos o datos y así sucesivamente. | | **~~Figura 7~~**  Mes de beneficio del préstamo, con frecuencia descendente.    ***~~Nota.~~*** *~~Muestra el mes del beneficio del préstamo con su frecuencia de forma descendente~~****~~.~~***  **Imagen:** 228131\_i246 |
| **Otras escalas de medición son la dicotómica y la cronológica**   * **La escala de medición dicotómica:** consta de dos categorías excluyentes.   **Ejemplo**: si el paciente posee una enfermedad, las respuestas son 2: sí o no.   * **La escala de medición cronológica:** los sucesos se ubican en una línea de tiempo.   **Ejemplo**: las acciones de Ecopetrol en la bolsa de valores de New York*,* durante el mes de octubre 2022. | | 10 aÃ±os escritos en un papel con una chincheta azul, imagen conceptual para visiÃ³n empresarial o prospectiva a largo plazo.  **Imagen:** 228131\_i289 |
| **Últimos datos de la población**  Se tiene una población de 12152 con 12 columnas.  En la figura se pueden observar las últimas 5 filas del data frame con sus respectivos índices**:** 12147, 12148, 12419, 12150, 12151. | | **~~Figura 8~~**  Últimos datos de la población.    ***~~Nota.~~*** *~~Muestra los últimos datos de la población.~~*  **Imagen:** 228131\_i247 |
| **Muestra aleatoria**  Se puede tomar una muestra aleatoria de un tamaño predefinido. Por ejemplo: 15 elementos del total de la población de 12152.  En la figura se pueden observar, los 15 elementos de la muestra tomados de forma aleatoria, con sus respectivos índices: 1104, 5844, 517, 10560, 11598, 1406, etc**.** | | **~~Figura 9~~**  Muestra aleatoria de 15 elementos de la población de 12152.    ***~~Nota~~****~~. Muestra aleatoria de 15 elementos de la población de 1215.~~*  **Imagen:** 228131\_i248 |

|  |  |  |
| --- | --- | --- |
| **Tipo de recurso** | Infografía interactiva 1 Punto caliente | |
| **Texto introductorio** | A continuación, analice la muestra aleatoria de un tamaño predefinido de 15 elementos: | |
| **~~Figura 10~~** *~~Muestra aleatoria de 15 elementos de la población~~* | | |
| **Código de la imagen** | 228131\_i249 | |
| **Punto caliente 1** | Muestra aleatoria de la población de un tamaño de 15 elementos. | P1 en amarillo. |
| **Punto caliente 2** | Cada uno de los 15 elementos escogidos al azar, enumerados por los índices 1104, 5844 ,517 de una población de 12152 individuos o elementos. | P2 en amarillo. |

|  |  |  |
| --- | --- | --- |
| **Tipo de recurso** | Infografía interactiva 2 Punto caliente | |
| **Texto introductorio** | Ahora, observe la muestra aleatoria de un 10 % de la población: | |
| **~~Figura 11~~** *~~Muestra aleatoria de un 10% de la población~~* | | |
| **Código de la imagen** | 228131\_i250 | |
| **Punto modal 1** | Muestra aleatoria de un 10% de la población. | Punto 1 en amarillo. |
| **Punto modal 2** | Cada uno de los 12152 elementos escogidos al azar, correspondientes al 10 % de la población de 12150. Enumerados por los índices 4172, 11471, 10863. | Punto 2 en amarillo. |

**Muestreo aleatorio estratificado**

|  |  |  |
| --- | --- | --- |
| **Tipo de recurso** | Slider Presentación | |
| **Introducción** | Este tipo de muestreo se utiliza cuando se quiere saber la relación entre los subgrupos o estratos, por alguna característica o múltiples características en común. Por ejemplo: el año en que fue otorgado el crédito, la edad de la persona que recibe el crédito, etc. | |
| De esta población de 12152, podemos dividirlo por estratos según el año 2017, 2018, 2019. La característica o estrato sería el año del préstamo. Tomando un subgrupo o subpoblación por año. Decidir el tamaño de la muestra de acuerdo a cada estrato. | | Imagen: 228131\_i251 |
| Se debe realizar el muestreo aleatorio en cada estrato. Por ejemplo, tomar los que fueron beneficiados en el 2018, que son 8765 y, de allí tomar la muestra aleatoria con un tamaño representativo determinado.  Analice la siguiente tabla:  **Tabla 1** *Años versus beneficiarios del crédito*   |  |  | | --- | --- | | Años de los beneficiarios del crédito | Número de beneficiarios del crédito | | 2017 | 1921 | | 2018 | 8765 | | 2019 | 1466 |   ***~~Nota.~~*** *~~Muestra los años y el número de beneficiarios del crédito de consumo.~~* | | Imagen: 228131\_i252 |

# Diseño de instrumentos, recolección de datos, verificación de la calidad

|  |  |
| --- | --- |
| **Tipo de recurso** | Acordeón tipo 1 |
| **Introducción** | En el diseño de instrumentos de recolección de datos, los métodos de recopilación de estos, se definen como un conjunto de procedimientos y herramientas para recopilar, validar y analizar la información necesaria para lograr los objetivos.  A continuación, conocerá algunos instrumentos que se utilizan en la estadística básica para cumplir con este propósito: |
| Imagen: 228131\_i253 | |
| **Diseño de instrumentos**  En la construcción formal de recolección de datos, generalmente se utiliza un cuestionario.  La elaboración del instrumento consta de dos etapas: la primera una etapa cualitativa, donde están las preguntas o ítems y la segunda etapa, es la evaluación o medición de las métricas de las preguntas o ítems. | |
| **Recolección de datos**  Puede ser mediante entrevista, encuesta, cuestionario, observación, diagrama de flujo, diccionario de datos. | |
| **La encuesta**  Consiste en aplicar un cuestionario a un grupo de personas, con la finalidad de sacar información sobre un tópico en particular. Hay dos tipos de encuesta: la analítica y la descriptiva. | |
| * **La encuesta analítica** busca describir y explicar una situación en particular. * **La encuesta descriptiva busca** documentar una situación en particular. | |
| **Cuestionario**  Un cuestionario es un documento, con un conjunto de preguntas que deben estar organizadas, secuenciadas y estructuradas para sacar la mayor información de las respuestas. | |

|  |  |
| --- | --- |
| **Tipo de recurso** | Acordeón tipo 2 |
| **Introducción** | A continuación, se encuentran algunos ejemplos de aplicación de instrumentos de recolección, utilizados en estadística básica. |
| Imagen general que ilustre el tema    Imagen: 228131\_i254 | |
| **Ejemplo de cuestionario:**   1. Pueden ser con preguntas abiertas, como:   ¿Qué opinión le merece la participación de los ciclistas colombianos en los diferentes tours internacionales?   1. Preguntas cerradas:   De selección única:  ¿Tiene apartamento propio? Las respuestas: Sí o No.  Pregunta con escala de valores:  En una escala del 1 a 10 siendo 1 el mínimo, ¿recomendaría nuestro producto? | |
| **Ejemplo de recolección de datos:**  Se quiere vender un producto conocido en un negocio de la localidad se necesita saber:  Se debe averiguar la edad de la persona a encuestar, el género, si consume el producto, con qué frecuencia consume el producto, si tiene hora fija para consumir el producto, si los miembros de la familia consumen el producto, si el hábito de compra del producto ha sido influenciado por las promociones en redes sociales. Además, indagar sobre el promedio mensual (cuánto gasta en el producto), la condición de salud del entrevistado. Asimismo, se plantea una pregunta abierta sobre el producto, qué necesitaría para cambiar de marca y el orden de importancia de otros productos referente al del estudio, etc. | |
| **Ejemplo de preguntas para un cuestionario:**  Se quiere hacer una encuesta para un negocio de venta en línea, posibles preguntas:  En una escala del 1 al 10 ¿recomendaría nuestro comercio a un amigo o familiar?  En una escala del 1 al 10 con base a su experiencia de usuario, ¿qué tan satisfecho se siente?   * ¿Cómo se enteró de nuestro negocio? * ¿Compra para otra persona? * ¿Regresaría a nuestro comercio? * ¿Es fácil navegar por el sitio? * Experiencia en la compra. * Experiencia en la llegada del producto. * ¿Los productos están bien explicados y detallados? * ¿Los productos tienen fotos de buena calidad? * ¿Los precios de los productos son similares a los del mercado? | |

**~~Verificación de la calidad:~~**

|  |  |
| --- | --- |
| **Tipo de recurso** | Tarjetas Conectadas |
| **Introducción** | Medir el nivel de confiabilidad, validez y calidad de los datos, permite encontrar e identificar errores que pueden corregirse y valorar, si estos son adecuados, para la finalidad requerida. |
| Imagen: 228131\_i255 | |
| Imagen: 228131\_i256 | **Confiabilidad y validez**  Para la verificación de la calidad se debe tener confiabilidad y validez. La confiabilidad indica que la aplicación repetida del instrumento al mismo individuo, arroje resultados idénticos. La validez se refiere al grado en el cual un instrumento mide aquello para lo cual fue creado. |
| Imagen: 228131\_i257 | **Evitar preguntas ambiguas**  Para la confiabilidad se deben evitar preguntas ambiguas, el instrumento debe abordar las variables de estudio y la medición podría cambiar a lo largo del tiempo. |
| Imagen: 228131\_i258 | **Medir la confiabilidad**  El **Alfa de Cronbach** es un índice que se utiliza para medir la confiabilidad, para saber que tanto están relacionados los ítems en un instrumento. Está basado en el promedio de las correlaciones de los ítems. |

**Ejemplo encuesta Invamer**

|  |  |
| --- | --- |
| **Tipo de recurso** | Acordeón tipo 1 |
| **Introducción** | Para ampliar y consolidar los aprendizajes, se muestra a continuación un ejemplo que ilustra el uso de los instrumentos de recolección estadística, mediante una empresa colombiana dedicada a la investigación y asesoría de mercado: |
| Imagen general que ilustre el tema    Imagen: 228131\_i259 | |
| **Encuesta Invamer ~~resultó~~**  El 34 % de los colombianos consideran que el desempleo y la economía, son los mayores problemas del país(Infobae, 2022 y El espectador, 2022).  Empresa que realizó la encuesta: Invamer S.A.S.  Fuente de financiamiento: recursos propios de Invamer S.A.S.  Pregunta la encuesta **Invamer Poll***,* que le preguntó a los entrevistados: ¿cuál es el mayor problema que tiene el país?  ~~A continuación, se observa una~~ **~~ficha técnica~~** ~~de una encuesta Invamer.~~ | |
| **Universo**  Hombres y mujeres de 18 o más años, de todos los niveles socioeconómicos, residentes en: Bogotá (5 479 011), Medellín (1.813.155), Cali (1.374.126), Barranquilla (810.019) y Bucaramanga (390.351), para un total de 9.866.662 personas, según Censo 2018. | |
| **Marco muestral**  Cartografía de las ciudades para seleccionar a las personas. El cubrimiento de la cartografía en las 5 grandes ciudades es del 95 %. La cantidad de manzanas en la cartografía son: Bogotá 39.103, Medellín 13.329, Cali 13.707, Barranquilla 7.783 y Bucaramanga 4.716 para un total de 78.638. | |
| **Tamaño y distribución de la muestra**  1200 encuestas personales distribuidas de la siguiente manera: Bogotá 400 encuestas, Medellín 200 encuestas, Cali 200 encuestas, Barranquilla 200 encuestas y Bucaramanga 200 encuestas. Para ajustar la muestra total al tamaño real del universo de cada ciudad, se aplican factores de ponderación. | |
| **Sistema de muestreo**  Se llevó a cabo un muestreo probabilístico por etapas, primero se realizó una selección aleatoria sistemática de hogares y posteriormente, una selección aleatoria simple de una persona de 18 años o más. | |
| **Margen de error**  Los márgenes de error dentro de unos límites de confianza de un 95 %, son: para el total de la muestra de las 5 ciudades +/- 2,83 %; para el total de la muestra de Bogotá +/- 4,90 %; para los totales de las muestras de Medellín, Cali, Barranquilla y Bucaramanga +/-6,93 %. | |
| **Técnica de recolección de datos**  Encuestas personales en el hogar de los encuestados (CAPI).   * Fecha de recolección de los datos: julio de 2022. * Número de encuestadores: 25 | |
| **Método de validación**  Se revisó el 100 % de las encuestas realizadas y se supervisó el 15 % de las mismas. | |
| **Temas a los que se refiere**  Opinión pública sobre gobernantes, personajes, instituciones y hechos de actualidad. | |
| **Factor de ponderación**  Se calculó dividiendo el porcentaje del universo de las variables, rango de edad y nivel socioeconómico, sobre el porcentaje de la muestra para las mismas variables. | |

**Análisis de la ficha técnica**

|  |  |  |
| --- | --- | --- |
| **Tipo de recurso** | Slider Presentación | |
| **Introducción** | Los documentos técnicos son documentos que contienen métodos y procesos para realizar estudios sociales o de mercado. La ficha técnica es un elemento muy útil de un estudio, debido a que, sirve para presentar la información requerida para poder interpretar los resultados de forma correcta.  Siguiendo con el caso de la encuesta Invamer, se presenta un ejemplo de análisis de ficha técnica que servirá de guía para la aplicación como analista de datos: | |
| Trabajan en cinco ciudades grandes Bogotá, Medellín, Cali, Barranquilla y Bucaramanga. Tamaño de la muestra 1200 personas, 400 en Bogotá por ser más grande, el resto 200. Límite del intervalo de confianza de 95 %, significa que hay un error tipo 1 ( ) del 5 %. Muestreo probabilístico por etapas. Encuestas personales en el hogar. | | Imagen: 228131\_i260 |
| Al utilizar la calculadora para obtener el tamaño de la muestra con un intervalo de confianza del 95 % **(QuestionPro, 2022)**. Se obtiene un tamaño de 384 para valores superiores a una población 90000 y menores a 220 999. Para valores de una población mayores a 220 999 se obtiene un tamaño de la muestra de 385. | | Imagen: 228131\_i261 |
| Según el DANE, Bogotá tiene al año 2022 una población aproximada de 7,9 millones de personas, si se pone esa cifra en la calculadora para obtener el tamaño de la muestra, arroja para un 5 % de error (en la casilla margen de error se escribe el número 5) y un intervalo de confianza del 95 %, un tamaño de muestra de 385. Se puede obtener el mismo tamaño de muestra para diferentes tamaños de población. En Invamer en Bogotá se aproxima el tamaño de la muestra a 400. | | Imagen: 228131\_i262 |

**Introducción a Google Colab**

|  |  |  |
| --- | --- | --- |
| **Tipo de recurso** | Slider Presentación | |
| **Introducción** | En Google Colab se tienen máquinas virtuales con tarjetas gráficas que Google facilita. Además, permite ejecutar código en el lenguaje Python, dentro de la nube, con el formato de Jupyter notebook extensión ipynb.  Jupyter notebook se puede utilizar localmente o en la nube, con Google Colab. Para usarlo localmente, se debe tener instalado Google Colab, 2022, en el computador. | |
| **Entorno**  Dentro de Google Colab, se está en un entorno interactivo denominado Notebook de Colab, que permite escribir y ejecutar código celda por celda, o todo de una vez; se obtiene en la pestaña Código.  Escribe en un lenguaje etiquetado especial llamado Markdown, se obtiene en la pestaña de nombre Texto**.** | | **~~Figura 12~~**  Pantalla de inicio Google Colab.    ***~~Nota.~~*** *~~Pantalla de inicio de Google Colab.~~*    Imagen: 228131\_i263 |
| **Carga de documentos**  En la figura,en el menú vertical de la izquierda, se observa un ícono o figura de una carpeta; haciendo clic allí, se pueden observar los documentos que se tiene y subir documentos a Google Colab, haciendo clic a una flecha de subida de documentos. | | LA MISMA IMAGEN DEL SLIDE ANTERIOR |
| **Instalar una librería o paquete**  **Ejemplo:**  pip install pandas  Instalaría la librería de nombre pandas, si ya está instalada lo dirá.  **Importar librería con un alias(as)**  **Ejemplo:**  import pandas as pd  Cargar la librería pandas, con el alias o abreviación pd.  **Ejemplo:**  import math  *Cargar la librería de nombre math referente a matemática.*  print(dir(math))  Muestra las funciones que están dentro de la librería math  pow(2,3) una de las funciones que están en la librería math, es la función potencia en inglés Power, pow(2,3) significa que eleva el número 2 a la 3 y da como resultado 8. | | **~~Figura 13~~**  Pantalla de inicio GC. Importar la librería math y utilizar la función potencia: pow().    ***~~Nota~~****~~.~~**~~Importar la librería math y utilizar la función potencia: pow().~~*  Imagen: 228131\_i264 |
| **Data frame**  Un Data frame es una estructura de datos con dos dimensiones (filas y columnas), se pueden guardar datos de distintos tipos (caracteres, enteros, valores de punto flotante, etc.) en columnas. Es similar a una hoja de cálculo o una tabla de SQL. | | Imagen: 228131\_i265 |
| **CSV**  Un archivo CSV (valores separados por comas) es un archivo de texto que tiene un formato específico, que permite guardar los datos en un formato de tabla estructurada.  Un archivo CSV puede tener diferentes separadores, distintos de la coma ( , ) como: punto y coma ( ; ), barra oblicua ( / ), guión del medio( - ), guion de piso ( \_ ), etc. | | Imagen: 228131\_i266 |
| **Leer un archivo CSV**  Para leer un archivo en formato CSV se utiliza el comando pd.read\_csv('nombre del archivo.csv'), pd es la abreviación de la librería pandas.  En la parte nombre del archivo, se debe escribir la ruta o dirección donde está el archivo, al final de esa ruta el nombre con extensión.csv. | | **~~Figura 14~~**  Primeras cinco filas del Data Frame de créditos de consumo otorgados.    ***~~Nota.~~*** *~~Muestra las primeras 5 filas del data frame de créditos de consumo otorgados.~~*  Imagen: 228131\_i267 |
| **Ubicación del archivo CSV**  Se puede utilizar el comando pwd, para saber la ruta donde se está ubicado.  Para cambiar el directorio o ruta se utiliza el comando cd.  **Ejemplo: cd ruta** del directorio.  **Ejemplo:** leer un archivo csv, utilizando el alias de pandas pd, nombrarlo df y visualizarlo con el nombre df, separados por punto y coma ( ; ) y ver las primeras 5 filas.  **import pandas as pd**  **df = pd.read\_csv('creditos\_de\_consumo\_otorgados.csv',sep=’;’)**  **df.head()** | | ICONO de archivo CSV con portátil. Tipo de documento de hoja de cálculo. Ilustración gráfica moderna de diseño plano. Icono CSV vectorial. - arte vectorial de Internet libre de derechos  Imagen: 228131\_i286 |
| df.tail()  Muestra las últimas 5 filas del data frame, de créditos de consumo otorgados. | | ***~~Figura 15~~***  Últimas cinco filas del data frame de créditos de consumo otorgados.    ***~~Nota~~****~~. Muestra las últimas 5 filas del data frame de créditos de consumo otorgados.~~*  Imagen: 228131\_i268 |
| **Información de variables**  Se puede saber información sobre las variables, utilizando en la sentencia el nombre del data frame seguido de .info(), también si las variables tienen valores únicos y cuántos son, utilizando el nombre del data frame seguido de .nuinque()  **Ejemplo**: df.info()  df.nunique() | | **~~Figura 16~~**Tipos de datos.    Muestra los tipos de datos y valores únicos del data frame de créditos de consumo otorgados.  Imagen: 228131\_i269 |
| **Estadísticas de una columna numérica**  Se puede obtener el valor mínimo, valor máximo, número de elementos de las filas, cuartil 1(Q1), cuartil 2(Q2), cuartil 3 (Q3), la media (mean), desviación estándar (std).  Se coloca el nombre de la columna, seguido de un punto y la palabra describe()  **Ejemplo:**  df['edad'].describe() | | **~~Figura 17~~**  Estadísticas básicas columna edad.    ***~~Nota.~~*** *~~Muestra las estadísticas básicas de la columna edad del data frame de créditos de consumo otorgados~~*~~.~~    Imagen: 228131\_i270 |
| Reemplazar el valor de una columna  Se puede reemplazar masculino, por ejemplo, con la letra M. Femenino, por ejemplo, con la letra F.    **Ejemplo:**  df['sexo']=df['sexo'].replace('masculino','M')  df['sexo']=df['sexo'].replace('femenino','F') | | **~~Figura 18~~**  Reemplazar en la columna sexo.    Muestra como remplazar en la columna sexo masculino por M y femenino por F del data frame de créditos de consumo otorgados.  Imagen: 228131\_i271 |
| **Selección aleatoria del 10 % de la población**  Para extraer una muestra aleatoria del 10 %, se importa la libreríarandom, se utiliza el nombre del dataframe punto sample(frac=0.1), 0.1 significa el 10 %.  En la figura se puede observar de la población 12152 el 10 % del tamaño de la muestra tomados de forma aleatoria con sus respectivos índices: 4172, 11471. El tamaño de la muestra es 1215.  **Ejemplo:**  **Importar la librería**  Import random  pob\_10 = df.sample(frac =.1)  pob\_10.head() | | **~~Figura 19~~**  Muestra aleatoria 10 %.    ***~~Nota.~~***  *~~Muestra una muestra aleatoria con el 10% de la población del data frame de créditos de consumo otorgados~~*~~.~~  Imagen: 228131\_i272 |
| **Muestra aleatoria**  Para extraer una muestra de tamaño 10, se importa la librería random y se utiliza el nombre de la data frame, seguido por .sample(n=10).  **Ejemplo:**  **Importar la librería**  Import random  muestreo10 = df.sample(n=10)  En la figura se puede observar de la población 12152 los 10 elementos de la muestra tomados de forma aleatoria con sus respectivos índices: 9263, 1335, 10100, 6998. | | **~~Figura 20~~**Muestra aleatoria.    Muestra una muestra aleatoria de 10 elementos de la población 12152 del data frame de créditos de consumo otorgados.  Imagen: 228131\_i273 |
| **Descarga del documento**  El formato de Jupyter notebook tiene extensión. ipynb. Dentro de Google Colab, se puede descargar con esa extensión o con la extensión del lenguaje python .py.  Como muestra la figura, haciendo clic en el menú horizontal superior, en la pestaña archivo, se baja hasta la pestaña descargar, allí se obtienen dos opciones: descargar en formato .ipynb de Jupyter Notebook y en formato .py corresponde al lenguaje Python*.* | | **~~Figura 21~~**  Forma de descarga documento.    Muestra la forma de descargar el documento Jupyter Notebook con extensión .ipynb o .py.  Imagen: 228131\_i274 |

|  |
| --- |
| Cuadro de texto |
| ~~Para obtener más información sobre el tema, se puede acceder a los siguientes enlaces TOOMEY, D. (2018). Learning Jupyter 5: Explore Interactive Computing Using Python, Java, JavaScript, R, Julia, and JupyterLab, 2nd Edition. Birmingham, UK: Packt Publishing.~~  ~~TOOMEY, D. (2017) Jupyter for Data Science : Your One-stop Guide to Building an Efficient Data Science Pipeline Using Jupyter. Birmingham: Packt Publishing.~~  ~~Imagen: 28131\_i275~~  ~~Que encuentran en material complementario~~ |

|  |
| --- |
| Cuadro de texto |
| Finalice su estudio de este componente, profundizando en los conceptos y procedimientos generales y claves de la estadística básica. Par ello, visualice el video que se propone a continuación: Reproducción de vídeo en la tableta  **VER ANEXO: 228131\_v02 – estadística básica**  Imagen: 28131\_i276  SE ENCUENTRA EN LA CARPETA CF02\_228131 – ESTADISTICA BÁSICA |

|  |
| --- |
| ~~Cuadro de texto~~ |
| ~~Para una mejor comprensión del video, el manejo de~~ *~~Google Colab~~*~~, es importante descargar los 3 archivos adjuntos:~~   |  |  |  |  | | --- | --- | --- | --- | | **~~Tema~~** | **~~Tipo de archivo~~** | **~~Nombre de archivo~~** | **~~Enlace/botón de descarga~~** | | ~~Manejo de~~ *~~Google Colab~~* | ~~.PDF~~ | ~~Manejo de Google Colab.pdf~~ |  | | ~~Créditos de consumo~~ | ~~.CSV~~ | ~~Creditos\_de\_consumo\_otorgados.csv~~ |  | | ~~Gráficos~~ *~~Ploty~~* | ~~.IPYNB~~ | ~~Estadística\_básica.ipynb~~ |  |   ~~SE ENCUENTRAN EN LA CARPETA CF02\_228131-Descargar Material~~ |

|  |
| --- |
| Cuadro de texto |
| Recuerde explorar los demás recursos que se encuentran disponibles en este componente formativo; para ello, diríjase al menú principal, donde encontrará la síntesis, una actividad didáctica para reforzar los conceptos estudiados, material complementario, entre otros. |

**Síntesis**

|  |  |
| --- | --- |
| **Tipo de recurso** | Síntesis |
| ~~Nombre del mapa: Estadística básica Síntesis:~~ | |
| **Introducción** | El siguiente mapa integra los criterios y especificidades de los conocimientos expuestos en el presente componente formativo: |
| **~~Figura 22~~** *~~Mapa conceptual estadística básica~~*    Imagen: 228131\_i277 | |

**Actividad interactiva**

|  |  |  |
| --- | --- | --- |
| **Tipo de recurso** | Actividad didáctica. Verdadero y falso | |
| Apreciado aprendiz, a continuación, encontrará una serie de preguntas que deberá resolver, con el objetivo de evaluar la aprehensión de los conocimientos expuestos en este componente formativo: | | Imagen: 228131\_i278 |
| 1. Una submuestra es tomar una muestra de la población. | | Imagen: 228131\_i279 |
| **Verdadero** | | Falso (correcto) |
| Retroalimentación:    Para obtener una submuestra, primero se debe obtener una muestra de la población y, a esa muestra sacarle otra muestra. | | |
| 1. En el muestreo aleatorio simple no todos los individuos tienen la misma probabilidad de salir en la muestra. | | Imagen: 228131\_i280 |
| **Verdadero** | | Falso (correcto) |
| Retroalimentación:  Es una técnica de muestreo, donde todos los individuos que conforman la muestra tienen la misma probabilidad de ser elegidos. Si el individuo puede ser seleccionado más de una vez, se está en presencia de un muestreo aleatorio simple con reposición; en caso contrario, sería un muestreo aleatorio simple sin reposición. | | |
| 1. En el muestreo aleatorio estratificado no se pueden distinguir subgrupos. | | Imagen: 228131\_i281 |
| **Verdadero** | | Falso (correcto) |
| Retroalimentación:  Este tipo de muestreo se utiliza cuando se quiere saber la relación entre los subgrupos o estratos, por alguna característica o múltiples características en común. Por ejemplo: el año en que fue otorgado un crédito, la edad de la persona que recibe el crédito, etc.  Ejemplo: se puede dividir la muestra de las personas a quienes les fue otorgado un crédito de consumo en 7 subgrupos por edad, cada letra corresponde a un grupo: a) 18-27 años, b) 28-37 años, c) 38-47 años, d) 48-57 años, e) 58-67 años, f) 68-77 años, g) 78-88 años. | | |
| 1. En el muestreo sistemático la población no está ordenada. | | Imagen: 228131\_i282 |
| **Verdadero** | | Falso (correcto) |
| Retroalimentación:  En el muestreo sistemático la población, al sacar la muestra, debe estar ordenada de alguna forma.  Ejemplo: Se va a extraer una muestra de 10 personas de una muestra de 50 personas. El primer individuo seleccionado para la muestra es el número 5 con un tamaño o intervalo de 4, decidido por el investigador.  Los elementos de la muestra serían: 5, 9, 13, 17, 21, 25, 29, 33, 37, 41. | | |
| 1. Un dato categórico ordinal representa categorías sin clasificación. | | Ilustración del concepto de tablero scrum  Imagen: 228131\_i283 |
| **Verdadero** | | Falso (correcto) |
| Retroalimentación:  En un dato categórico ordinal, los valores representan categorías con alguna clasificación intrínseca, tiene características de orden. Por ejemplo: malo, bueno, excelente. | | |
| 1. Un dato categórico nominal representa categorías sin clasificación intrínseca. | | Ilustración del concepto de tablero scrum  Imagen: 228131\_i284 |
| **Verdadero (correcto)** | | Falso |
| Retroalimentación:  Un dato categórico nominal, representa categorías sin clasificación intrínseca. Por ejemplo: las marcas de zapatos, género, raza, etnia. | | |
| 1. La escala de medición dicotómica consta de dos categorías. | | Escala de nivel de metabolismo de infografía plana con flecha y valor de medición  Imagen: 228131\_i285 |
| **Verdadero (correcto)** | | Falso |
| Retroalimentación:  La escala de medición dicotómica consta de dos categorías excluyentes. Por ejemplo: si el paciente posee una enfermedad, las respuestas son dos: si o no. | | |
| 1. La elaboración del instrumento consta de dos etapas. | | Lista gigante  Imagen: 228131\_i286 |
| **Verdadero (correcto)** | | Falso |
| Retroalimentación:  La elaboración del instrumento tiene dos etapas. La primera es una etapa cualitativa donde están las preguntas o ítems y, la segunda etapa es la evaluación o medición de las métricas de las preguntas o ítems. | | |
| 1. La recolección de datos puede ser mediante entrevistas. | | Grupo profesional trabajando en proyecto.  Imagen: 228131\_i287 |
| **Verdadero (correcto)** | | Falso |
| Retroalimentación:  La recolección de datos puede ser mediante entrevista, encuesta, cuestionario, observación, diagrama de flujo, diccionario de datos. | | |
| 1. Una encuesta es aplicar un cuestionario a un grupo de personas. | | Ilustración del concepto de exámenes  Imagen: 228131\_i288 |
| **Verdadero (correcto)** | | Falso |
| Retroalimentación:  Una encuesta es aplicar un cuestionario a un grupo de personas con la finalidad de sacar información sobre un tópico en particular. Hay dos tipos de encuesta la analítica y la descriptiva.  La encuesta analítica busca describir y explicar una situación en particular y la encuesta descriptiva busca documentar una situación en particular.  Un cuestionario es un documento con un conjunto de preguntas deben estar organizadas, secuenciales y estructuradas, para sacar la mayor información de las respuestas. | | |

**Material complementario**

|  |  |  |  |
| --- | --- | --- | --- |
| Tipo de recurso | Material complementario | | |
| Tema | Referencia APA del material | Tipo | Enlace |
| **Estadística básica** | Aguilar, J. (2021). *Estadística Descriptiva regresión y probabilidades con aplicaciones*. Ediciones de la U. | Libro digital | <https://www-ebooks7-24-com.bdigital.sena.edu.co/?il=15156&pg=202> |
| **Diseño de instrumentos, recolección de datos, verificación de la calidad** | Moreno, A. & Corcoles, S. (2020). *Python práctico*. Ediciones de la U. | Libro digital | <https://www-ebooks7-24-com.bdigital.sena.edu.co/?il=10285&pg=1> |
| **Estadística básica** | Triolla, M. (2018). *Estadística* (12 ed.) Pearson Education. | Libro digital | <https://www-ebooks7-24-com.bdigital.sena.edu.co/?il=7386&pg=149> |
| **Diseño de instrumentos, recolección de datos, verificación de la calidad** | Toomey, D. (2018). *Learning Jupyter 5: Explore Interactive Computing Using Python, Java, JavaScript, R, Julia, and JupyterLab* (2nd Edition). Packt Publishing. | Libro digital | <https://search-ebscohost-com.bdigital.sena.edu.co/login.aspx?direct=true&db=e000xww&AN=1881491&lang=es&site=ehost-live> |
| **Diseño de instrumentos, recolección de datos, verificación de la calidad** | Toomey, D. (2017) *Jupyter for Data Science: Your One-stop Guide to Building an Efficient Data Science Pipeline Using Jupyter*. Packt Publishing. | Libro digital | <https://search-ebscohost-com.bdigital.sena.edu.co/login.aspx?direct=true&db=e000xww&AN=1637911&lang=es&site=ehost-live>. |

**Glosario**

|  |  |
| --- | --- |
| **Tipo de recurso** | Glosario |
| Censo | Un conjunto de actividades destinadas a recoger, recopilar, el padrón o lista de la población de una nación o pueblo en un momento determinado, la información demográfica social y económica, para luego evaluar, analizar y publicar o divulgar. |
| Continua | Una variable continua es aquella que puede tomar un número infinito de valores entre dos valores cualesquiera de una característica. La altura de los 5 persona: 1.85, 1.89, 1.93, 1.59, 1.65. En la práctica medimos la altura con dos decimales, pero también se podría dar con tres decimales o más. |
| Cualitativa | Son aquellas en la que los resultados posibles no son valores numéricos. Por ejemplo: color del pelo, tipo de ropa preferida, lugar de veraneo, etc. |
| Cuantitativa | Se trata de características cuyas variaciones pueden ser expresadas de forma numérica. Por ejemplo: edad, peso, estatura, número de hijos, años de servicio, etc. |
| Cuartiles | Son 4 divide un conjunto de datos ordenados en cuatro partes iguales. Cuartil1 (Q1) 25 %, la mediana es el Cuartil 2 (Q2) 50 %, Cuartil 3 (Q3) 75 % Cuartil 4 (Q4) 100 % se trabaja principalmente con los Q1, Q2, Q3. |
| Dato atípico | Es un valor mayor a 1.5 veces el rango intercuartil. |
| Desviación estándar | Mide la dispersión de una distribución de datos. Entre más dispersa está una distribución de datos, más grande es su desviación estándar. |
| Discreta | Una variable es discreta cuando no puede tomar ningún valor entre dos consecutivos, y que es continua cuando puede tomar cualquier valor dentro de un intervalo. Ejemplos de variable discreta: número de empleados de una fábrica; número de hijos; número de cuentas en un banco. |
| Escala Cronológica | El orden cronológico de los sucesos es la ubicación de estos en una línea temporal. |
| Escala nominal | Es una escala de medición en la cual los números sirven como “etiquetas” solamente para identificar o clasificar un objeto. Una escala de medición nominal normalmente trata sólo con variables no numéricas (no cuantitativas). |
| Escala ordinal | Una escala ordinal u “ordenada” permite evaluar la actitud de un encuestado con respecto a un tema utilizando, un conjunto de respuestas ordenadas. Por ejemplo, las respuestas pueden incluir: “Muy satisfecho”, “Satisfecho”, “Insatisfecho” y “Muy insatisfecho”. |
| Estadísticos | Es cualquier función real medible de la muestra de una variable aleatoria. |
| Media | Resultado que se obtiene al dividir la suma de varias cantidades por el número de sumandos. |
| Muestreo | Selección de un conjunto de personas o cosas, que se consideran representativos del grupo al que pertenecen, con la finalidad de estudiar o determinar las características del grupo. |
| Rango intercuartil | Es una medida de dispersión. Se obtiene restando el cuartil 3 (Q3) menos el cuartil 1(Q1); Q3-Q1 en ese rango está el 50 % de los datos. |

**Referentes bibliográficos**

|  |  |
| --- | --- |
| **Tipo de recurso** | Bibliografía |
| Altair. (2022). *Vega-Altair: Visualización declarativa en Python*. <https://altair-viz.github.io/> | |
| :  E~~l Espectador. (2022).~~ *~~Repunte de optimismo en Colombia tras elecciones presidenciales~~*~~. El espectador.~~ [~~https://www.elespectador.com/politica/repunte-de-optimismo-en-colombia-tras-elecciones-presidenciales/~~](https://www.elespectador.com/politica/repunte-de-optimismo-en-colombia-tras-elecciones-presidenciales/) | |
| *Google Colaboratory.* (s. f.). *Te damos la bienvenida a Colab.* <https://colab.research.google.com/?hl=es> | |
| Infobae. (022). *Encuesta Invamer: el 34 % de los colombianos considera que el desempleo y la economía son los mayores problemas del país*. <https://acortar.link/UzYzDT> | |
| Lind, D. & Marchal, W. (2019). *Estadística Aplicada a los negocios y la economía* (17 ed.). Mc Graw Hill. <https://www-ebooks7-24-com.bdigital.sena.edu.co/?il=9246&pg=315> | |
| Matplotlib. (2022). *Matplotlib: visualización con Python.* <https://matplotlib.org/> | |
| Martínez, C. (2019). *Estadística y muestreo* (14 ed.). Ecoe ediciones. <https://www-ebooks7-24-com.bdigital.sena.edu.co/?il=9278&pg=387> | |
| Martínez, C. (2019). *Estadística básica aplicada* (5ta ed.). Ecoe ediciones. <https://www-ebooks7-24-com.bdigital.sena.edu.co/?il=8975&pg=82> | |
| Metadata Alcaldía de Medellín. (2022). <https://acortar.link/zGU2gK> | |
| Seaborn. (2022). *Seaborn: visualización de datos estadísticos.* <https://seaborn.pydata.org/> | |
| Plotly. (2022). *Plotly Biblioteca de gráficos de código abierto para Python.* <https://plotly.com/python/> | |
| Questionpro. (2022). *Calculadora de Muestras.* <https://www.questionpro.com/es/calculadora-de-muestra.html> | |
| Vladimirovna, O. & Gutiérrez, E. (2016). *Estadística inferencial 1 para ingeniería y ciencias.* Grupo Editorial Patria. Recuperado de: <https://elibro-net.bdigital.sena.edu.co/es/ereader/senavirtual/40474?page=69> | |





