**Datos de identificación del programa de formación**

| PROGRAMA DE FORMACIÓN | Manejo de la información para sistemas de gestión |
| --- | --- |

| COMPETENCIA | 260101067. Procesar los datos de acuerdo con técnica de investigación y tabulación. | RESULTADOS DE APRENDIZAJE | 260101067-01. Diligenciar instrumentos para la recolección de información según el Sistema de Gestión de la organización.  260101067-02. Seleccionar los datos basados en los métodos de estadística.  260101067-03. Ordenar los datos de acuerdo con las técnicas y normas de investigación.  260101067-04. Tabular datos de acuerdo con el sistema de gestión de la organización.  260101067-05. Reportar datos procesados según sistemas de información y codificación vigente. |
| --- | --- | --- | --- |

| NÚMERO DEL COMPONENTE FORMATIVO | 1 |
| --- | --- |
| NOMBRE DEL COMPONENTE FORMATIVO | Recolección de información y bases de datos |
| BREVE DESCRIPCIÓN | La recolección de datos ha cobrado gran importancia en la actualidad gracias a su utilidad para anticipar tendencias a través del análisis de la información obtenida por distintas fuentes o métodos. Se hace necesario comprender lo relacionado a este proceso, su relación con las bases de datos y su contribución en entornos organizaciones. |
| PALABRAS CLAVE | Base de datos, tabulación, estadística, procesamiento. |

| ÁREA OCUPACIONAL | 1 - FINANZAS Y ADMINISTRACIÓN |
| --- | --- |
| IDIOMA | Español |

**TABLA DE CONTENIDOS**

**Introducción**

**1. La recolección de datos**

**2. La estadística**

2.1. Métodos estadísticos y técnicas de muestreo

2.2. Medidas de dispersión y tipos de variables

2.3. Estructura de datos

**3. ¿Qué es una base de datos?**

3.1. Estructura y elementos de las bases de datos

3.2. La tabulación

**4. Herramientas de trabajo**

**INTRODUCCIÓN**

| Cuadro de texto |
| --- |
| Apreciado aprendiz en este componente se familiarizará con los conceptos relacionados a la ciencia de datos y conocerá los procesos para su identificación, recolección, tabulación y análisis en diversos entornos, en especial los organizacionales. De esta forma, se pretende establecer condiciones para la comprensión de su significado, y se abordará la importancia que tiene el uso del manejo de bases de datos en procesos donde se deba tomar decisiones. Para conocer acerca de este componente lo invitamos a ver el siguiente video: |

| **Tipo de recurso** | Video animación 2D | | | |
| --- | --- | --- | --- | --- |
| **Título** | Introducción | | | |
| **Escena** | **Imagen** | **Sonido** | **Narración (voz en off)** | **Texto** |
| **Escena 1** |  | Música ambientando el video. | La evolución de la sociedad se ha dado a causa de cuatro revoluciones, estas han aportado complejidad e incertidumbre al contexto de las organizaciones. La última de ellas es llamada ‘revolución 4.0’ que llevó a la sociedad al estadio actual, esta se caracteriza por el desarrollo tecnológico, y por los cambios vertiginosos en aspectos económicos, sociales, demográficos, ambientales, entre otros. Por ende, se presenta una amplia disponibilidad y posibilidades de acceso a la información en el entorno, hecho que establece una serie de necesidades y desafíos para las organizaciones en los procesos de análisis y toma de decisiones. | Revolución 4.0  Información |
|  |  |  | La dinámica del desarrollo tecnológico ha abierto las puertas a un sinfín de posibilidades, las cuales están relacionadas con la generación y obtención de datos valiosos para el progreso de las organizaciones. Es así, como el uso de diferentes herramientas tecnológicas pueden asegurar en la organización sostenibilidad, crecimiento y efectividad en los procesos de interacción con sus grupos de interés. Sin embargo, este proceso es cada vez más complejo porque requiere identificar fuentes de información de mayor idoneidad y pertinencia. | Obtención de datos  Herramientas tecnológicas |
|  |  |  | Por otra parte, la consolidación de la denominada ‘transformación digital’ implica la necesidad de apropiar métodos y herramientas diseñadas para garantizar idoneidad en la captación, procesamiento, análisis y uso de datos relacionados con los grupos de interés de las organizaciones. De esta forma, se asegura a la gerencia una mayor eficacia en la toma de decisiones, y posterior comunicación a nivel interno y externo. | Captación, procesamiento y uso de datos |
|  |  |  | En resumen, este componente formativo tiene como objetivos conocer las técnicas para la recolección, tabulación y análisis de datos y comprender la importancia de estos para la toma de decisiones dentro en las organizaciones. | Recolección, tabulación y análisis |
|  |  |  | Para ello, se abarcarán conceptos relacionados con ciencia de datos y estadística, además se darán a conocer las herramientas y procedimientos para recolectar, procesar y analizar datos con enfoque organizacional. |  |
|  |  |  | Inicialmente, se abordarán los conceptos básicos, técnicas, ejemplos y herramientas relacionadas con la identificación y recolección de datos. | Conceptos y herramientas |
|  |  |  | En segundo lugar, se presentarán metodologías de gran utilidad para el desarrollo de procesos de identificación, recolección, procesamiento y análisis de datos. | Metodologías |
|  |  |  | Adicionalmente, se realizará la caracterización, utilidad, construcción y aplicabilidad de las bases de datos en el contexto organizacional, y los mecanismos para su gestión, actualización y depuración. | Caracterización  Utilidad  Aplicabilidad |
|  |  |  | Agregado a lo anterior, se explorará el concepto de sistemas y metodologías de flujo de trabajo (w*ork flow).* Esta herramienta es de gran utilidad porque genera eficiencia en los procesos, haciendo énfasis en aquellos relacionados con la gestión de la información. | Sistemas y metodologías  *Work flow* |
|  |  |  | Además, es necesario abordar diferentes técnicas que permiten tabular, presentar y analizar series de datos fundamentales para la toma de decisiones en las organizaciones. Este análisis se realiza a través de la apropiación e implementación de diferentes herramientas tecnológicas. | Técnicas de tabulación  Herramientas tecnológicas |
|  |  |  | Al culminar el proceso, tendrá los conocimientos necesarios para realizar procesos de análisis de información sensible en las organizaciones de forma eficaz y eficiente, y estará en la capacidad de brindar una mayor rigurosidad y fluidez en la toma de decisiones operativas y estratégicas. | Conocimientos para brindar mayor fluidez |

**DESARROLLO DE CONTENIDOS**

1. **La recolección de datos**

| Cuadro de texto |
| --- |
| Para definir el significado de la recolección de datos es necesario precisar qué un dato según la RAE es “información sobre algo concreto que permite su conocimiento exacto o sirve para deducir las consecuencias derivadas de un hecho”. De acuerdo a lo anterior, se puede definir a la recolección de datos como la acción de reunir información sobre algo concreto. A continuación se conocerán otras definiciones relacionadas con la recolección de datos: |

| **Tipo de recurso** | Slider Presentación | |
| --- | --- | --- |
| **Introducción** |  | |
| El diccionario de la Universidad de Cambridge establece que los datos son “información, especialmente hechos o números, recopilados para ser examinados, considerados y utilizados para ayudar en la toma de decisiones; o información en forma electrónica que puede ser almacenada y utilizada por un computador”. Por lo anterior, se puede concluir que la palabra ‘dato’ está relacionada con el procesamiento al que se puede someter y extraer información para la tomar decisiones. | | **Imagen:** 135401\_i1 |
| Los datos se consideran uno de los activos más importantes para las organizaciones, es frecuente que quien tenga la información tiene el poder, puesto que la información es la base de todo proceso que se quiera realizar en el entorno. El desarrollo de las tecnologías de la información (TIC) han facilitado el acceso a gran cantidad de datos; acción que genera la necesidad de gestionar adecuadamente la información, y en especial la denominada ‘sensible’.  Gómez, I et al. (2018), en su libro *Calidad de Datos* indica que “la transformación digital está afectando a todos los sectores económicos y sociales. Convirtiendo a los datos en la base para la toma de decisiones y el habilitador más potente para cualquier organización”. (p. 25). Es por ello que las organizaciones deben estar a la vanguardia tecnológica, teniendo como objetivo el uso adecuado y seguro de la información presentada en el entorno. | | **Imagen:** 135401\_i2 |
| Para Harari, Y. (2016), en *Homo Deus. Breve historia del mañana*, concibe el denominado ‘dataísmo’ como “una religión en la que el valor de cualquier fenómeno o entidad se determina por su capacidad para contribuir al procesamiento de datos”. (p. 438), hecho que se ve reflejado en diversos escenarios donde se requiere formalizar procesos para la identificación, recolección y análisis de datos con base en la estructura de la estrategia organizacional. | | **Imagen:** 135401\_i3 |

**¿Para qué sirve la recolección de datos?**

| Cuadro de texto |
| --- |
| Después de conocer algunos planteamientos de lo que significa la recolección de datos se revisará para qué son útiles. Para ello, Dunstan, J. et al. (2022), en *Una mirada a la era de los datos,* plantean que “el dato es una representación parcial de un fenómeno, que puede ser registrada, transmitida, almacenada y procesada a fin de obtener información del mismo”. (p. 14)  **Los datos abren la puerta al conocimiento,** es por ello que el proceso de identificación, recolección y análisis se ven determinados por la disponibilidad de información y la diversidad de fuentes de consulta. También, es necesario contar con la divulgación de información no verificable, ya que a causa de esta se deben establecer mecanismos que permitan dar respuesta a las siguientes preguntas:   1. ¿Qué datos se requieren? 2. ¿De qué fuentes obtenerlos? 3. ¿Qué instrumentos de recolección se deben utilizar? 4. ¿Cómo procesarlos? 5. ¿Cómo analizarlos? |

| Cajón de texto de color |
| --- |
| Es importante tener en cuenta que la calidad del análisis de los datos dependerá de la rigurosidad con que se desarrolle el proceso, de igual manera, es necesario comprender que los datos tienen un ciclo de vida y que de su comprensión dependerá en gran medida del resultado del análisis que se realice.  **Figura 1** *Ciclo de vida del dato*      Nota. El ciclo de vida del dato presenta la ruta que transitan los datos desde su generación hasta su interpretación. Adaptado de Wing, J. (p. 2), 2019. *The Data Life Cycle*. <https://hdsr.mitpress.mit.edu/pub/577rq08d/release/3>. |

| Cuadro de texto |
| --- |
| Por otro lado, algunas de las ventajas que ofrece el proceso de recolección de datos son las siguientes: |

| **Tipo de recurso** | Carrusel de tarjetas | |
| --- | --- | --- |
| **Introducción** |  | |
| Facilitar el análisis de datos y comprender el contexto en que fueron obtenidos. | | **Imagen:** 135401\_i4 |
| Definir los criterios de organización que contribuyen a la formulación de estrategias para cada grupo de interés. | | **Imagen:** 135401\_i5 |
| Identificar las oportunidades para la organización. | | **Imagen:**135401\_i6 |
| Reconocer las debilidades en los procesos internos y formular acciones de mejora. | | **Imagen:** 135401\_i7 |
| Analizar y comprender el comportamiento de los clientes. | | **Imagen:** 135401\_i8 |

**Técnica y herramientas de recolección de datos**

| Cuadro de texto |
| --- |
| Como se ha visto, la recolección de información permite acceder a los datos, y se logra a través de técnicas e instrumentos cuya selección obedece a aspectos como el objetivo de la recolección, tiempo requerido, recursos disponibles, conocimiento previo, entre otros.  A continuación se presentan algunas de las técnicas utilizadas con mayor frecuencia para la recolección de datos: |

| **Tipo de recurso** | | Pestañas o tabs Verticales |
| --- | --- | --- |
| **Imagen:** 135401\_i9 | | |
| **Observación** | Es una técnica simple que pretende reconocer el comportamiento del objeto de estudio de manera directa, y su aplicación no requiere gran disponibilidad de recursos. Un observador puede realizar el proceso partiendo de la definición de las variables a observar y registrándolas en una bitácora. Para realizar esta técnica se puede hacer uso de una herramienta digital así se agiliza el proceso. | |
| **Cuestionarios o encuestas** | Es un mecanismo de recolección popular que permite obtener datos directamente de los sujetos de interés. Para garantizar de forma correcta su diseño es necesario tener en cuenta las variables requeridas y formular preguntas que permitan cumplir los objetivos. Esta técnica puede incluir preguntas abiertas o cerradas dependiendo de los datos requeridos; la construcción de las preguntas y las opciones de respuesta obedecerán a la calidad de los datos obtenidos. Es por ello que un cuestionario o encuesta debe garantizar un proceso ágil de recolección. | |
| **Grupos focales** | Consiste en desarrollar una entrevista colectiva que permite contrastar las respuestas de los participantes y fomentar la discusión sobre el tema tratado, de esta manera es posible enriquecer los datos y ver las diversas perspectivas sobre una problemática. Su efectividad depende de la capacidad del moderador para controlar la discusión y evitar que se desvié o se convierta en un debate; además él será quien genere un ambiente de confianza entre los participantes y observe aspectos concernientes al lenguaje no verbal y a las reacciones de los participantes. | |
| **Entrevistas** | A diferencia del grupo focal, las entrevistas son un tipo de herramienta que se realiza de manera individual, y consiste en la aplicación de un cuestionario que puede ser estructurado o no. Antes de aplicarse se debe definir las variables a identificar para obtener los datos adecuados. Esta herramienta requiere de empatía e interacción con el entrevistado, así crear un ambiente cómodo que certifique el éxito en su aplicación.  El entrevistador debe garantizar que la conversación con el entrevistado sea amena y fluida para que no se convierta en una indagatoria, deberá estar atento a las manifestaciones del lenguaje no verbal (gestos, acciones, posturas, etc.), ya que de ese diálogo depende la calidad de los datos que se obtengan. | |

**Herramientas para la recolección de datos**

| Cuadro de texto |
| --- |
| Para realizar el proceso de recolección de datos hay varias herramientas que se encuentran disponibles y dependen del tipo de información, cantidad y el objetivo buscado. Estas herramientas se comparten a continuación: |

| **Tipo de recurso** | Acordeón Tipo 1 |
| --- | --- |
| **Introducción** |  |
| **Imagen:** 135401\_i10 | |
| **Reconocimiento Óptico de Caracteres (OCR)**  Este método consiste en digitalizar información permitiendo identificar el texto y hacer uso de este a través de herramientas de procesamiento de texto; es aplicado cuando se dispone de grandes cantidades de información de texto impresa o en formato de imagen. | |
| **Procesamiento Inteligente de Documentos (IDP)**  Permite establecer una estructura de datos extraídos desde archivos históricos no estructurados, facilitando la toma de datos, mitigando el desgaste operativo y generando plantillas de datos aptos para el análisis. Esta herramienta basa su solución en el uso de la Inteligencia Artificial (AI). | |
| **Almacén Operacional de Datos (ODS)**  Es un reporteador enlazado a los sistemas de información organizacional el cual garantiza que la información sea centralizada y presentada de acuerdo con las necesidades de análisis y toma de decisiones para los usuarios con acceso. | |

**2. La estadística**

| Cuadro de texto |
| --- |
| La estadística es una disciplina de orden matemático, se centra en **la interpretación de datos cuantitativos**; permite analizar fenómenos reales o experimentales a través de la recolección y ordenamiento de series de datos y así facilitar los procesos de toma de decisiones. |

| Tipo de recurso | Cajón de texto de color |
| --- | --- |
| Peña, D. (2015) en *Fundamentos de estadística*, plantea que “la estadística actual es el resultado de la unión de dos disciplinas que evolucionan independientemente hasta confluir en el siglo XIX: la primera es el cálculo de probabilidades, que nace en el siglo XVII como teoría matemática de los juegos de azar; la segunda es la «estadística» (o ciencia del Estado, del latín Status), que estudia la descripción de datos y tiene unas raíces más antiguas. La integración de ambas líneas de pensamiento da lugar a una ciencia que estudia cómo obtener conclusiones de la investigación empírica mediante el uso de modelos matemáticos.” (p. 21). | |

| Tipo de recurso | Cajón de texto de color |
| --- | --- |
| Además plantea el autor que “La estadística actúa como disciplina puente entre los modelos matemáticos y los fenómenos reales. Un modelo matemático es una abstracción simplificada de una realidad más compleja, y siempre existirá cierta discrepancia entre lo observado y lo previsto por el modelo. La estadística proporciona una metodología para evaluar y juzgar estas discrepancias entre la realidad y la teoría”. Es posible dimensionar la importancia de esta disciplina en el contexto actual caracterizado por la masificación de datos y los requerimientos de análisis para la toma de decisiones. | |

**Tipos de estadística**

| Cuadro de texto |
| --- |
| La estadística se considera una valiosa herramienta de organización, análisis, comprensión y fluidez para los procesos que conlleven la toma de decisiones, siendo útil y adaptable en el entorno organizacional. Esta disciplina se divide en dos ramas que son la estadística descriptiva y la estadística inferencial que se describen a continuación: |

| **Tipo de recurso** | Rutas / Pasos. Verticales 2 | |
| --- | --- | --- |
| **Introducción** |  | |
| **a** | **Estadística descriptiva**  Según Bologna, E. (2018), la estadística descriptiva provee una serie de procedimientos dirigidos a resumir, sintetizar y volver manejable la información para poder interpretarla y extraer conclusiones a partir del conjunto de datos que, de otra manera, serían ininteligibles. (Métodos estadísticos de investigación*,* p. 15). | Imagen: 135401\_i11 |
| **b** | **Estadística inferencial**  Bologna, E. (2018), indica que esta estadística proporciona los procedimientos para estudiar un subconjunto de elementos y generalizar las conclusiones obtenidas. Esta estadística analiza unas variables en una muestra, y estima, por inferencia, los valores de la población. Con la aplicación de ella se reducen costos, y disminuye el tiempo necesario para adelantar investigaciones; además, si se respetan los procedimientos de la estadística inferencial el error puede ser conocido y controlado.(p. 15) | Imagen: 135401\_i12 |

**2.1. Métodos estadísticos y técnicas de muestreo**

| Cuadro de texto |
| --- |
| Para comenzar a estudiar sobre los métodos estadísticos es importante conocer que el análisis estadístico permite realizar descripciones respecto a los valores de datos de diverso origen (económicos, políticos, sociales, organizacionales, entre otros). Esta herramienta permite predecir, relacionar y analizar la información aportada por estos datos.  **La información permite controlar, optimizar, administrar, examinar, investigar, planificar, predecir o tomar decisiones**, y lograrlo requiere de mecanismos para su desarrollo que la estadística los aporta. La efectividad del proceso de análisis y toma de decisiones depende de la disposición suficiente de datos actuales e históricos y, dependiendo del objetivo de análisis, es posible hacer uso de alguno de los siguientes seis métodos: |

| **Tipo de recurso** | Acordeón tipo 1 |
| --- | --- |
| **Introducción** | Métodos estadísticos |
| **Imagen:** 135401\_i13 | |
| **Análisis descriptivo**  Permite hacer descripciones respecto a un conjunto de datos con el fin de distinguir sus las características. Este método aporta el análisis de variables como el número de ventas realizadas, cantidad de clientes nuevos, unidades de producto disponibles en el inventario, entre otros. | |
| **Análisis exploratorio**  Reconoce una comprensión básica de los datos al identificar características sobresalientes, inesperadas o con valores atípicos. Este tipo de análisis permite determinar las sucursales con un número superior de ventas, productos con mayor *stock*, principales discrepancias entre el producto más vendido y el menos vendido, diferencias en aspectos económicos entre dos poblaciones, entre otros. Este método es un proceso descriptivo, por ende debe realizarse sin ideas preconcebidas sobre la información que aportan los datos. De este modo, en aras de evitar el procesamiento de datos inapropiados, erróneos o inadvertidos lo más aconsejable es que este análisis exploratorio sea utilizado durante la etapa inicial de todo análisis de datos. | |
| **Análisis inferencial**  Confirma o niega hipótesis planteadas permitiendo establecer conclusiones con cierto nivel de confianza, este tipo de métodos no garantizan certeza absoluta porque conservan algún grado de incertidumbre. Este análisis permite formular hipótesis basadas en conclusiones construidas respecto a un conjunto de datos en un entorno específico, por ejemplo: inferir que los clientes que adquieren determinado producto pertenecen a una familia numerosa (lo que los haría clientes potenciales para productos complementarios). Con el fin de garantizar efectividad en el análisis inferencial los datos deben ser diferentes a los empleados para la formulación de la hipótesis así se eliminan sesgos y se aporta validez a las conclusiones. | |
| **Análisis predictivo**  Identifica relaciones entre las variables en eventos pasados, así realizar proyecciones sobre posibles resultados en ocasiones futuras. Este análisis hace posible determinar la cantidad de producto que se debe tener en inventario por referencia o proyectar las principales necesidades de una población específica para los escenarios futuros; además se preocupa por el valor, no tiene como objetivo comprender el sistema o la relación entre variables, haciendo que el análisis sea funcional para estudiar problemas complejos y de amplio número de variables influyentes o de difícil comprensión. | |
| **Análisis causal**  Este tipo de análisis permite establecer relaciones de causa y efecto, identificar la influencia mutua de variables, predecir y reconocer la causa de este tipo de influencia. Además, este análisis determina aspectos como la afectación de las ventas en los productos de un portafolio respecto al día que aumentan la ventas del producto estrella, pero no las causas de este fenómeno; en otras palabras, el análisis inferencial permite identificar la relación, pero para conocer las causas es necesario implementar un análisis causal. | |
| **Análisis mecanicista**  Predice y comprende la influencia entre diferentes variables siendo este análisis el más complejo porque determina cómo la alteración de los valores de una de las variables tiene impacto en las demás. Su complejidad lo hace poco utilizado debido a que requiere amplio conocimiento de los procesos a modelar. Un ejemplo de aplicabilidad de este análisis es conocer si el incremento en la inversión en publicidad y un aumento de la plantilla en el departamento de marketing influye en el incremento de ventas y cómo esta última influye en las dos anteriores. | |

**Técnicas de muestreo**

| Cuadro de texto |
| --- |
| Para entender un poco más la disciplina de la estadística se debe conocer acerca del muestreo y sus técnicas. El muestreo es un conjunto de procedimientos para la selección de una determinada población y su subconjunto que se conoce como muestra; con el fin de conocer sobre la población escogida se observa, analiza y generaliza las características encontradas y provenientes de la muestra evaluada.  A continuación podrá detallar las diferentes técnicas de muestreo que se pueden utilizar: |

| **Tipo de recurso** | Acordeón tipo 1 |
| --- | --- |
| **Texto introductorio** |  |
| **Imagen:** 135401\_i14 | |
| **Muestreos probabilísticos**  Permiten generalizar los resultados obtenidos de la muestra a toda la población de referencia. Para que una muestra sea probabilística sus elementos deben ser elegidos aleatoriamente sin que el investigador decida a quién incluir o excluir de la muestra. | |
| **Muestreo aleatorio simple**  Consiste en realizar un sorteo para seleccionar aleatoriamente a los individuos que participarán en el proceso de recolección de información. | |
| **Muestreo sistemático**  Dividir la población en listas proporcionales, y seleccionar un número determinado de individuos de cada lista de manera aleatoria hasta completar la muestra. | |
| **Muestreo estratificado**  Consiste en estratificar la población en grupos, de acuerdo con criterios específicos, formando una relación de manera aleatoria hasta completar el tamaño de muestra. | |
| **Muestreo en conglomerado**  Se aplica cuando el tamaño de la población es muy grande y no es posible elaborar un listado de los individuos que la componen. Este muestreo requiere información estadística que estime su tamaño (ej. Último censo poblacional), información sobre la distribución por grupos poblacionales o zonas geográficas, y después seleccionar, de manera aleatoria, los miembros de la muestra. | |
| **Muestreos no probabilísticos**  En estas muestras no se cumple el requisito de aleatoriedad en la selección de los elementos que la componen. Los resultados no se pueden generalizar de manera probabilística más allá de los casos observados. (Bologna, E., 2018).  Este muestreo tiene definida una intencionalidad con los miembros de la muestra seleccionados, así el investigador tiene la posibilidad de establecer los criterios que deben cumplir los participantes. | |
| **Muestreo por conveniencia**  En este muestreo el investigador aplica el principio de oportunidad y selecciona a los miembros de su muestra de acuerdo con la facilidad que tenga para acceder a ellos. | |
| **Muestreo por cuotas**  A través de este tipo de muestreo el investigador establece criterios equitativos y proporcionales para la selección de los individuos que serán seleccionados para ser parte de la muestra. | |
| **Muestreo bola de nieve**  Este muestreo permite al investigador solicitar apoyo de los participantes para identificar a los miembros de la muestra, y se da cuando la selección de la muestra debe cumplir con unas características tan específicas que no son rastreadas fácilmente. | |
| **Muestreo discrecional**  Esta técnica permite la conformación de un grupo determinado para participar como miembros de la muestra, esta obedece a características específicas y definidas por el investigador, él será quien se encargue de seleccionar con discreción las personas que considere más adecuadas para ser incluidas en la muestra. | |

**2.2. Medidas de dispersión y tipos de variables**

| Cuadro de texto |
| --- |
| Luego de conocer las diferentes técnicas de muestreo, entraremos a analizar otro aspecto importante: **las medidas de dispersión**.  Ramos, F. & Guerra, R (2019), definen las medidas de dispersión como “un rasgo principal de los datos es su dispersión o amplitud, que se refiere a su variabilidad, a la evaluación de cuán separados o extendidos están estos datos, o bien cuánto difieren unos de otros”. (*Introducción a los métodos estadísticos*, p. 45). Otro concepto clave en las medidas de dispersión es ‘la variación’ que es el grado en que los datos numéricos tienden a extenderse alrededor de un valor, generalmente el valor medio. (Ramos, F. & Guerra, R (2019). Para ampliar más información sobre la variación revise el documento **Variación y medidas de dispersión.**  Para tener claridad en las medidas de dispersión, a continuación se mostrarán tres de las más utilizadas: |

| **Tipo de recurso** | Tarjetas Avatar |
| --- | --- |
| **Introducción** |  |
| **Rango o recorrido (Amplitud)**  Es la distancia entre los valores máximo y mínimo, y mide la dispersión de la totalidad de los datos. | |
| **La varianza (S)**  Media de las desviaciones cuadráticas entre cada valor de la variable y la media aritmética de la distribución. | |
| **Desviación típica**  Cuando se utiliza la varianza como medida de dispersión surge el problema de tener que trabajar con distintas dimensiones, por este motivo se define la desviación típica como la raíz cuadrada de la varianza. (Ramos, F. & Guerra, R., 2019). | |

**Variables cuantitativas y cualitativas**

| Cuadro de texto |
| --- |
| Considerando que la naturaleza de las observaciones es de gran importancia a la hora de elegir el método estadístico más apropiado para abordar su análisis, es fundamental tener claro cómo se clasifican las variables. Haciendo énfasis en lo anterior, a continuación conocerá los tipos de variables y sus características: |

| **Tipo de recurso** | Pestañas o tabs horizontales | |
| --- | --- | --- |
| **Introducción** |  | |
| **Variables cuantitativas o numéricas** | Son las variables que pueden medirse, cuantificarse o expresarse numéricamente, y pueden clasificarse en dos tipos:  **- Variables cuantitativas discretas** no admiten todos los valores intermedios en un rango, pues suelen tomar solamente valores enteros (número de hijos, etc.). Los datos que se obtienen para variables cuantitativas discretas se cuentan.  **- Variables cuantitativas continuas** admiten tomar cualquier valor dentro de un intervalo numérico determinado (longitud, peso, etc.). Los datos que se obtienen para variables cuantitativas continuas se miden a través de una escala de intervalo o de una escala proporcional (de razón). | Imagen: 135401\_i15 |
| **Variables cualitativas** | Representan una cualidad o atributo que clasifica a cada caso en una de varias categorías. Sus valores no se asocian naturalmente a un número sino a una la situación más sencilla que clasifica cada caso en uno de dos grupos (hombre/mujer).  Son datos dicotómicos o binarios, en muchas ocasiones este tipo de clasificación no es suficiente y se requiere de un mayor número de categorías para su clasificación como nacionalidad, profesión, entre otros.  En el proceso de observación de estas variables se pueden utilizar dos escalas nominales u ordinarias, además estas pueden codificarse a través de números para facilitar su procesamiento, pero sin olvidar qué tipo de variables son.  Una de las escalas más utilizadas en las variables cualitativas es ‘la escala de Likert’, en ella se establece una codificación ordenada y frecuentemente del 1 al 5 o también se utiliza la escala de Likert modificada con la de ‘la escala Thurstone’ de 10 puntos. La selección de la escala es basada en lo que el investigador desee lograr y las comparaciones que quiera realizar en la propia investigación o con otros trabajos desarrollados anteriormente. (Ramos, F. & Guerra, R, 2019, p. 11). | Imagen: 135401\_i16 |

**2.3. Estructura de datos**

| Cuadro de texto |
| --- |
| La naturaleza de los datos, su origen y utilidad son de gran diversidad, por ello se han definido diferentes categorías o escalas de medición que se adaptan a las particularidades de los mismos, al requerimiento de organización y al análisis de quien dispone de ellos. Con el fin de comprender las características propias de la estructura de datos es importante explorar las diferentes escalas de medición que se disponen a nivel estadístico para tal fin, las cuales se presentan a continuación: |

| **Tipo de recurso** | Acordeón tipo 1 |
| --- | --- |
| **Introducción** |  |
| **Imagen:** 135401\_i17 | |
| **Escala nominal**  Las observaciones de un carácter nominal se pueden clasificar en varias categorías, excluyentes entre sí, entre las que no es posible establecer ninguna relación de orden u operar matemáticamente. En este tipo de escala se consideran los atributos categóricos como el sexo, estado civil, tipo de contrato laboral, lugar de nacimiento, sector de actividad económica, entre otros. | |
| **Escala ordinal**  Las observaciones en escala ordinal se pueden clasificar en varias categorías, excluyentes entre sí, en ellas es posible establecer alguna relación de orden, pero no operar matemáticamente. Además, estas categorías se pueden ordenar, pero no cuantificar la diferencia entre ellas. En este tipo de escala se encuentran atributos ordinales como el nivel de ingresos, nivel de estudios, grado de satisfacción, entre otros. | |
| **Escala de intervalo**  Las observaciones en escala de intervalo se dan cuando existe una unidad de medida que permite cuantificar la distancia existente entre dos observaciones, siendo el cero arbitrario.  En este tipo de escala se consideran algunas variables (caracteres cuantitativos) como la temperatura en grados Celsius o en grados Fahrenheit, fechas, entre otros. | |
| **Escala de razón (proporción o cociente)**  Las observaciones son de escala de razón cuando existe una unidad de medida que permite cuantificar la distancia existente entre dos observaciones y existe un cero absoluto. En este tipo de escala se encuentra la mayoría de las variables (caracteres cuantitativos) como el peso, salario, nivel de inventario, la edad, entre otros. | |

| Cuadro de texto |
| --- |
| Así como en la estructura de datos las escalas de medición permiten organizar los datos en orden jerárquico, es importante conocer los diferentes tipos de estructuración de datos, en los cuales se encuentran las tablas de frecuencia, los gráficos estadísticos y los gráficos para datos categóricos. A continuación se dará una breve explicación de cada uno de ellos: |

| **Tipo de recurso** | Slider Presentación | |
| --- | --- | --- |
| **Introducción** |  | |
| **Tablas de frecuencia:** cuando la muestra presenta un alto número de datos es necesario organizarlos en una tabla de frecuencia (distribución de frecuencias). Este tipo de tabla es un método para presentar la información resumida por distintas categorías, valores o intervalos de carácter x y, por las siguientes frecuencias: absoluta (f), relativa (h), absoluta acumulada (H) y relativa acumulada (F). Para conocer más acerca de los tipos de frecuencia puede revisar el documento **Tipos de frecuencia.** | | **Imagen:** 135401\_i18 |
| **Gráficos estadísticos**: son recursos visuales que permiten comunicar la información, se considera en un instrumento estadístico para analizar los datos en forma eficiente y desde diferentes niveles de lectura. | | **x** |
| **Gráficos para datos categóricos:** de acuerdo con el tipo de frecuencia se pueden encontrar gráficos que ayudan a visualizar y a categorizar los datos. En este tipo de esquemas se encuentran los que son de tipo barra, torta, arañas, barras acumuladas, pictogramas, anillos, entre otros. | | **Imagen:** 135401\_i20 |

| Tipo de recurso | Cajón de texto de color |
| --- | --- |
| Para construir una tabla de frecuencia se tienen varias herramientas y una de las más usadas es el programa Excel. Este programa permite construir tablas que facilita el conteo y proceso de construcción; ya que permite crear en otras celdas la frecuencia deseada con el uso fórmulas, además replica la información de forma rápida obteniendo resultados exactos y sencillos. | |

| Cuadro de texto |
| --- |
| A continuación podrá estudiar los tipos de gráficos para datos categóricos propuestos por Castellanos, M. & Obando, J. (2021), los cuales serán citados textualmente, (p. 13 – 16): |

| **Tipo de recurso** | Tarjetas Animadas |
| --- | --- |
| **Introducción** |  |
| **Imagen:** 135401\_i21 | **Gráfico de barras**  Permite representar los elementos de la variable de estudio con los valores obtenidos en la frecuencia absoluta o relativa. En el eje X están las variables y en el eje Y el valor de las frecuencias. |
| **Imagen:** 135401\_i22 | **Gráfico de sectores**  Denominado también “tartas” (*pie*), consiste en un círculo dividido en tantas partes o porciones como valores de frecuencia relativa existan, a cada valor de frecuencias relativa le corresponde en forma proporcional un arco de círculo.  La lectura de estos gráficos es directa, se lo hace teniendo en cuenta el valor del porcentaje propuesto en cada unidad de variable. Una interpretación más general estará amarrada al objetivo de la investigación que generó la pregunta. |
| **Imagen:** 135401\_i23 | **Gráfico de araña o radar**  Una gráfica de radar, también conocida como un “diagrama de araña”, es una herramienta muy útil para mostrar visualmente los valores de una frecuencia relativa o absoluta. Proporciona la misma información de un diagrama de barras o de una torta. Es una excelente alternativa para representar la información de los datos propuestos en las frecuencias relativas, brinda la misma información, pero desde otra visualización. |
| **Imagen:** 135401\_i24 | **Gráfico de anillos o sectores**  Otra de las formas de representar las frecuencias relativas en variables categorías lo constituyen los anillos o los sectores. Son pequeños sectores en forma de un aro o anillo, divido en forma proporcional de acuerdo con el valor de dicha frecuencia.  Los diagramas para las frecuencias absolutas acumuladas, denominados “diagramas integrales”, no tienen sentido para las variables categóricas, se observarán en el contexto de variables numéricas, enteras y continuas. |

**3. ¿Qué es una base de datos?**

| Cuadro de texto |
| --- |
| Una base de datos (BD) es un repositorio de información que se encuentra organizado de tal forma que permite al usuario acceder, gestionar y actualizar los datos de forma sencilla.. Para conocer más acerca de las bases de datos lo invitamos a ver el siguiente video: |

| **Tipo de recurso** | Video animación 2D | | | |
| --- | --- | --- | --- | --- |
| **Título** | Fundamentos para entender las bases de datos | | | |
| **Escena** | **Imagen** | **Sonido** | **Narración (voz en off)** | **Texto** |
| **1** |  | Música de fondo | ¿Qué son las bases de datos?  Las bases de datos son herramientas que ayudan a las diferentes áreas de la organización a través de la creación de sistemas de información que redunden en una mejor comunicación, y por ende, se convierten en una unidad de criterio para el análisis de información y la toma de decisiones. | ¿Qué es una base de datos?  Recopilación organizada y estructurada de información |
| **2** |  | Música de fondo | La validez de la información que se encuentra en las bases de datos está estrechamente relacionada con el cumplimiento de los siguientes criterios:   * Independencia de los datos: la información consignada en ella no depende de un programa específico y puede ser procesada haciendo uso de diferentes aplicaciones. * Reducción de la redundancia: se hace el mayor aprovechamiento del espacio de almacenamiento y debe evitar la duplicidad de información. * Seguridad: la información debe estar protegida frente a usuarios no autorizados. | Independencia de los datos  Reducción de la redundancia  Seguridad |
| **3** |  | Música de fondo | Adicionalmente, una base de datos debe cumplir con tres criterios fundamentales con el propósito de garantizar fiabilidad y oportunidad de la información:   * Integridad * Redundancia * Consistencia | Características de una base de datos |
| **4** |  | Música de fondo | Iniciaremos explicando la integridad, esta hace referencia a la consistencia y validez de los datos que se encuentren en la base de datos. Responde al cumplimiento de criterios inviolables que se aplican a los datos, y su relación con los usuarios que acceden a estas bases de datos o quienes las alimentan. | Integridad  Criterios inviolables |
| **5** |  | Música de fondo | Una adecuada gestión de la información velará por el cumplimiento de los criterios de integridad de los datos que reposan en las bases de datos, para ello, la alta gerencia debe definir criterios para la alimentación, uso y depuración de estas herramientas.  Es importante asegurarse que los datos permanezcan intactos y sin cambios desde la captura hasta la última etapa del ciclo de vida, de esta forma, es posible evitar información errónea que pueda conducir a sesgos en la toma de decisiones. | Se debe tener en cuenta la integridad de los datos y su mantenimiento. |
| **6** |  | Música de fondo | Como segundo criterio se encuentra la redundancia, esta hace referencia a la multiplicidad de datos; es decir al almacenamiento de un dato varias veces en diferentes lugares de la base de datos, convirtiéndose en un alto riesgo para la calidad de la información que en reposa en la base de datos.  Puede generar problemas de un mal aprovechamiento del espacio de almacenamiento, además de una inminente inconsistencia de los datos.  Con el fin de evitarla, y asegurar la pertinencia y viabilidad de la información para su análisis y posterior toma de decisiones, es necesario establecer criterios o mecanismos que permitan ingresar y depurar la información que se encuentre en la base de datos. | Redundancia  Desventajas:  Reducción de la capacidad de almacenamiento y dificultad en la búsqueda de información. |
| **7** |  | Música de fondo | Finalmente, se encuentra la consistencia la cual es entendida como la coherencia entre a la información de la base de datos. Este criterio asegura la pertinencia de los datos y la validez de estos en el proceso de la toma de decisiones.  En caso de inconsistencia en los datos el proceso de análisis y toma de decisiones podrá incurrir en sesgos que pueden afectar la sostenibilidad y el crecimiento de las organizaciones; ya que esta incongruencia podría llevar a tomar decisiones basadas en datos inexactos o incorrectos. | Cuando se pierde la integridad, se pierde la consistencia. |

**Tipos de bases de datos**

| Cuadro de texto |
| --- |
| Una vez abordados los conceptos más representativos en relación con las bases de datos, su funcionalidad y nivel de aporte para el entorno organizacional, es necesario contar con elementos de valor que permitan identificarlas, y sacar el máximo provecho de la información que estas contienen. A continuación, se presentan las tres categorías que definen cómo clasifican las bases de datos, esta clasificación no obedece a su nivel de importancia o utilidad, sino son definidas teniendo en cuenta el tipo de datos que en ellas se albergan y el objetivo de su uso. Los siguientes criterios son citados según Pulido et al.(2019, p. 20 - 21): |

| **Tipo de recurso** | Infografía interactiva Modal | |
| --- | --- | --- |
| **Texto introductorio** | Estos son los tres criterios en los que se clasifican las bases de datos: | |
| **Imagen**    La imagen es de referencia, ya que debe dar contexto de muchos datos, carpetas, folders, documentos, imágenes de gráficos estadísticos y deben aparecer en el siguiente orden:  1. Formulario como los que se llenan en supermercados y tiendas.  2. Una guía telefónica electrónica, junto a una biblioteca virtual (que se ven los títulos de los libros, los autores y las diferentes maneras en que presentan la información de los contenidos)  3. Documento donde se detalle el organigrama de una empresa, y una pantalla donde se vea un *ecommerce* | | |
| **Código de la imagen** | 135401\_i25 | |
| **Punto modal 1** | **Según la variabilidad:**   * **Estáticas:** son bases de datos diseñadas en exclusiva para lectura y la cual almacena datos históricos para analizar su comportamiento a través del tiempo para hacer proyecciones, tomar decisiones y realizar inteligencia empresarial. * **Dinámicas:** almacenan información que se puede modificar con el tiempo. Permiten actualizar, borrar y editar los datos, además de hacer operaciones fundamentales de consulta. Un ejemplo de este tipo de base de datos es la que utiliza el sistema de información de un supermercado. | En la imagen que tendrá el formulario |
| **Punto modal 2** | **Según el Contenido**   * **Bibliográficas:** contienen una fuente primaria que permite localizarla. Un registro típico de una base de datos bibliográfica contiene información sobre el autor, fecha de publicación, editorial, título, edición, de una determinada publicación, entre otras. Además pueden contener un resumen o extracto de la publicación original. * **De texto completo:** almacenan las fuentes primarias, como ejemplo se encuentran las guías telefónicas en formato electrónico. * **Bibliotecas de información:** son bases de datos que almacenan diferente tipo de información relacionada con disciplinas científicas como química, ciencias de la vida o medicina, mismas que a su vez se dividen en varios subtipos. (Pulido et al.2019, p. 21). | En la imagen donde se encuentre la guía telefónica electrónica. |
| **Punto modal 3** | **Según su modelo de base de datos**   * **Jerárquica:** los datos se organizan en forma de árbol invertido (conocido como raíz), en donde un nodo padre de información puede tener varios hijos. * **De red:** es un modelo similar al jerárquico pero se diferencia de este porque permite que un mismo nodo tenga varios padres, ofreciendo una solución eficiente al problema de la redundancia de datos. * **Transaccionales:** son bases de datos cuyo único fin es el envío y recepción de datos a grandes velocidades, estas bases son poco comunes y están dirigidas, por lo general, al entorno de análisis de calidad, producción o industrial. Su propósito es recolectar y recuperar datos a la mayor velocidad posible, en este modelo la redundancia o duplicación no representa un problema. * **Relacionales:** se utiliza para representar problemas reales y administrar datos de manera dinámica. Se fundamenta en el uso de relaciones que pueden considerarse en forma lógica como los conjuntos de datos llamados tuplas. * **Multidimensionales:** permiten desarrollar aplicaciones muy concretas como Cubos OLAP. Son muy similares a las bases de datos relacionales pero se diferencian de estas porque presentan campos o atributos de dos tipos, es decir, pueden representar las diferentes dimensiones de la tabla e incluso las métricas que se desea aprender. * **Orientada a objetos:** este modelo es de reciente creación y es propio de los modelos informáticos. Tiene como propósito almacenar en la base de datos los objetos completos en cuanto a su estado y comportamiento. * **Documentales:** consienten en la indexación de texto completo y en líneas generales. Este método permite realizar búsquedas más potentes ya que almacenan grandes volúmenes de información entre las cuales se encuentran registros históricos. * **Deductivas:** permite hacer deducciones a través de inferencias, y principalmente se basa en reglas y hechos almacenados en la base de datos. Son llamadas bases de datos lógicas porque se basan en la lógica matemática. | En la imagen donde aparezca el organigrama de una empresa y el *ecommerce*. |

**3.1. Estructura y elementos de las bases de datos**

| Cuadro de texto |
| --- |
| Agregado a lo anterior, las bases de datos pueden ser estructuradas en niveles arquitectónicos de tres o cuatro niveles, estos guardan estrecha relación con la manera en que se almacenan y articulan los datos, teniendo en cuenta la estructura definida, y el modo en que los datos interactúan con los usuarios que tienen acceso a la BD. |

| **Tipo de recurso** | Pestañas o tabs horizontales | |
| --- | --- | --- |
| **Introducción** | Las bases de datos pueden tener dos arquitecturas y serán citadas según lo expuesto por Pulido et al.(2019, p.22), a continuación: | |
| **Arquitectura de tres niveles** | 1. **Nivel interno o físico:** es el más cercano al almacenamiento físico. Describe los detalles de cómo se almacenan físicamente los datos: los archivos que contienen la información, su organización, los métodos de acceso a los registros, los tipos de registros, la longitud y los campos que los componen. 2. **Nivel externo o de visión:** es el más cercano a los usuarios. Cada esquema describe la parte de la BD que es de interés para un determinado grupo de usuarios, en este nivel se representa la visión individual de un usuario o de un grupo. 3. **Nivel conceptual**: describe la estructura de toda la BD para un grupo de usuarios mediante un esquema conceptual. Este esquema describe las entidades, atributos, relaciones y operaciones de los usuarios, así como las restricciones, ocultando los detalles de las estructuras físicas de almacenamiento. | **Imagen:** 135401\_i26 |
| **Arquitectura de cuatro niveles** | 1. **Nivel conceptual o visión**: se describe la información de la organización (objetos y relaciones) desde un punto de vista no informático. 2. **Nivel lógico o visión:** se representan las entidades y relaciones de acuerdo con las características de dicho modelo sin entrar en el espacio de la representación física. 3. **Nivel interno:** descripción de la representación en la memoria externa de la computadora de los datos del esquema lógico, sus interrelaciones y los instrumentos para acceder a ellos. 4. **Niveles externos:** cada uno describe los datos y relaciones que entre ellos son de interés para una aplicación dada. Estos esquemas pueden verse como subconjuntos del modelo lógico de la BD. | **Imagen:** 135401\_i27 |

| Cuadro de texto |
| --- |
| Como otro punto a tener en cuenta es que las bases de datos se componen por diversos elementos, entre los cuales se destacan cuatro que corresponden a tablas, formularios, informes y consultas; estos se explican a continuación: |

| **Tipo de recurso** | Infografía interactiva Modal | |
| --- | --- | --- |
| **Texto introductorio** | Elementos de una base de datos | |
| **Imagen**    La imagen debe contener los diferentes elementos de los que se habla en el contenido:   * Tablas de hoja de cálculo * Formulario * Hojas de informe * Una persona realizando una consulta (con una lupa) | | |
| **Código de la imagen** | **Imagen:** 135401\_i28 | |
| **Punto modal 1** | **Tablas**  Es similar a la apariencia de una hoja de cálculo, pues los datos se almacenan en filas y columnas. Para lograr la máxima flexibilidad la información debe estar organizada en tablas, para que no se tenga redundancia.  Cada fila de una tabla es un ‘registro’ donde se almacena cada información individual, este consta de campos (al menos uno). Los ‘campos’ corresponden a las columnas de la tabla y se deben configurar con un determinado tipo de dato, bien sea de texto, fecha, hora, numérico, o cualquier otro. | El primer punto va en el número 1 con la imagen de tablas de hojas de cálculo. |
|  | **Formularios**  Se conocen como las pantallas de entrada de datos. Los formularios se utilizan para trabajar con los datos y contienen botones que ejecutan diversos comandos. Se puede crear una base de datos sin usar formularios solo editando los datos de las tablas en las hojas. No obstante, es ideal usar formularios para ver, escribir y editar datos en las tablas. | El segundo punto va en el número 2 con la imagen del formulario. |
|  | **Informes**  Son ‘reportes’ que permiten resumir y presentar los datos de las tablas. Se pueden ejecutar en cualquier momento y siempre reflejarán los datos actualizados de la base de datos. En estos informes es posible consultar los datos en la pantalla, exportarlos en el computador o enviarlos por correo electrónico. | El tercer punto se encuentra el número 3 con la imagen de hojas de informe. |
|  | **Consultas**  Pueden realizar varias funciones, pero la más común es la de recuperar datos específicos de las tablas. Las consultas permiten agregar criterios para filtrar los datos hasta obtener los registros que desee. | El cuarto punto va donde se encuentra el número 4 con la imagen de una persona realizando una consulta (con una lupa) |

**3.2. La tabulación**

| Cuadro de texto |
| --- |
| Durante el procesamiento se usan los datos brutos como elemento de entrada una vez ingresados, estos deben converger para producir un elemento de salida como informes u otras herramientas de análisis.  A continuación, se indican los pasos a seguir para realizar el procesamiento de la información: |

| **Tipo de recurso** | Slider Hitos/ Línea de tiempo horizontal | |
| --- | --- | --- |
| **Introducción** | **Etapas de procesamiento de datos** | |
| **Recopilación de datos** | Este es el primer paso en el procesamiento de datos y consiste en extraen los datos de diferentes fuentes, pueden ser formatos de texto, almacenes de datos, tablas y matrices generadas por fuentes externas. | **Imagen:** 135401\_i29 |
| **Preparación de datos** | Los datos sin procesar deben ser depurados, organizados y verificados de manera minuciosa para identificar posibles errores, eliminando los datos incorrectos (redundantes, incompletos o incorrectos),de esta forma tener datos de alta calidad, y con una mayor inteligencia empresarial. | **Imagen:** 135401\_i30 |
| **Entrada de datos** | La entrada de datos es la primera etapa en que los datos sin procesar comienzan a convertirse en información utilizable, estos deben ser traducidos a un idioma comprensible. De este modo, los datos quedarán limpios y se ingresan a un destino específico que puede ser un CRM o un almacén de datos para su tratamiento. | **Imagen:** 135401\_i31 |
| **Procesamiento** | El procesamiento se puede realizar mediante técnicas de filtrado, análisis y visualización, e incluso con algoritmos de aprendizaje automático. El proceso puede variar dependiendo la fuente de los datos (base de datos, redes sociales, entre otros), y su uso previsto (examen de patrones de publicidad, determinar las necesidades del cliente, entre otros). | **Imagen:** 135401\_i32 |
| **Interpretación de los datos** | Es la etapa en que los datos son utilizables para los usuarios. Al traducir los datos se pueden leer en forma de gráficos, videos, imágenes, texto sin formato, entre otros.  En este punto es posible comenzar a administrar los datos para los proyectos de análisis de información. | **Imagen:** 135401\_i33 |
| **Almacenamiento de datos** | Una vez se procesan todos los datos son almacenados para su uso. Si bien, es posible que alguna información se use de inmediato o tenga un propósito más adelante. | **Imagen:** 135401\_i34 |

**Matrices**

| Cuadro de texto |
| --- |
| La matriz de datos permite ordenar un conjunto de observaciones dentro de un esquema de filas y columnas; es un instrumento que tiene una serie de elementos y puede tratarse de distintos objetos matemáticos y numéricos. Cada observación corresponde a la intersección de una fila con una columna, de este modo el ordenamiento permite apreciar aspectos como las características de una muestra o de una población estadística.  Una matriz está compuesta por tres elementos: variables, casos y valores; cada uno de ellos se detallará a continuación: |

| **Tipo de recurso** | | Pestañas o tabs Verticales |
| --- | --- | --- |
| **Introducción** | | Elementos de una matriz de datos |
| **Imagen:** 135401\_i35 | | |
| **Variables** | Son las características de una muestra o población, pueden ser cuantitativas o cualitativas. Las primeras son expresadas en números como la altura o el ingreso mensual, y las segundas indican una calidad como la profesión o el estado civil. | |
| **Casos** | Se encuentran en las filas y son elementos indivisibles que forman parte de un grupo sujeto a investigación. Sobre cada caso se analizan variables pertinentes al estudio realizado. Además, los casos pueden ser individuos, familias, grupos u otro, lo anterior dependerá de la estructura del trabajo estadístico. | |
| **Valores** | Es el objeto matemático que corresponde a la coincidencia de una variable y un caso. | |

**Técnicas de codificación**

| Cuadro de texto |
| --- |
| Guillermo, P. (s.f), en *Noción y estructura del dato* citando a Comboni, S. & Juárez, J. (*Introducción a las técnicas de investigación*), afirman que “La codificación es un procedimiento técnico mediante el cual, los datos obtenidos se clasifican en categorías y se traducen en símbolos, ya sean cifras o letras; es decir, se asigna a cada opción de respuestas un número o una letra que permita tabularla rápidamente.” (p. 5).  Dado lo anterior, se puede decir que la elaboración de un sistema de categorías y la codificación se puede hacer de forma simultánea, pues la codificación depende del sistema de categorías que adopte la variable o las alternativas que presente la pregunta. |

| **Tipo de recurso** | **Cajón de texto de color** |
| --- | --- |
| Asimismo, desde una perspectiva cualitativa, la codificación es un modo sistemático de desarrollar y refinar las interpretaciones de los datos. El proceso de codificación incluye la reunión y análisis de todos los datos que se refieren a temas, ideas, conceptos, interpretaciones y proposiciones. Durante esta etapa del análisis, lo que inicialmente fueron ideas e intuiciones vagas se refinan, expanden, descartan o desarrollan por completo. | |

| Cuadro de texto |
| --- |
| Para desarrollar ese sistema de categorías y codificación se pueden aplicar diversas técnicas. A continuación, se explicarán las cinco fases para codificar datos cualitativos: |

| **Tipo de recurso** | Rutas / Pasos. Verticales 1 |
| --- | --- |
| **Imagen:** 135401\_i36 | |
| **Botón 1** | Desarrollo de las categorías de codificación, empezando por redactar una lista de todos los temas, conceptos e interpretaciones y tipologías. |
| **Botón 2** | Codificación de todos los datos, escribiendo en el margen el número asignado o la letra correspondiente a cada categoría. |
| **Botón 3** | Separación de los datos según las diferentes categorías de codificación. |
| **Botón 4** | Revisión de qué datos han sobrado, estos se pueden ajustar a las categorías de codificación existentes o plantear unas nuevas. |
| **Botón 5** | Perfeccionamiento del análisis y comparación de los diferentes fragmentos relacionados con cada tema, concepto y propósito. |

| Cuadro de texto |
| --- |
| Una vez realizada la codificación de los datos se puede proceder a la confección de la matriz de datos, la cual no es más que el ordenamiento de los datos para su posterior análisis. |

**4. Herramientas de trabajo**

| Cuadro de texto |
| --- |
| Una vez abordadas las herramientas estadísticas para el análisis de datos y la toma de decisiones, es importante contar con instrumentos que aporten al desarrollo de las acciones encaminadas a la agrupación de los datos con el fin de garantizar su adecuado uso. De igual forma, se requiere el conocimiento de mecanismos que permitan dar fluidez a los procesos de trabajo, garantizando optimización de recursos y eficiencia en el procesamiento de los datos. |

**El sistema de flujo de trabajo**

| Cuadro de texto |
| --- |
| Debido a la amplia diversidad de los procesos de negocio y a las necesidades particulares que presentan las organizaciones, los flujos de trabajo son flexibles y se adaptan con facilidad a dichas características. A partir de su utilidad y los objetivos del proceso, es posible establecer y describir los tipos de *Work flow* expuestos por la Universidad Militar Nueva Granada (s.f) en *Sistemas de Work Flow*. (p. 5). |
|  |

| **Tipo de recurso** | Tarjetas Avatar | |
| --- | --- | --- |
| **Introducción** | Tipos de *Work Flow* | |
| ***Work flow* de producción**  Automatizan procesos de negocios de carácter repetitivo, estructurados y con gran número de datos. | | **Imagen** 135401\_i37 |
| ***Work flow* de colaboración**  Estructuran procesos de negocio en el que participan personas con un objetivo común. Involucran documentos denominados contenedores de información y siguen la ruta definida, además de las acciones que se toman sobre estos. | | **Imagen** 135401\_i38 |
| ***Work flow* de administración**  Involucra procesos de administración empresarial como órdenes de compra, reportes de ventas, entre otros. Se emplean cuando existe una gran cantidad de procesos administrativos dentro de la empresa. | | **Imagen** 135401\_i39 |

**La tabulación de datos**

| Tipo de recurso | Cajón de texto de color |
| --- | --- |
| Al momento de realizar un estudio es necesario determinar las condiciones del entorno en el cual se obtendrán datos, pues el entorno es muy grande, y es complejo abarcar todos los individuos que lo componen. La estadística plantea la posibilidad de definir un tamaño de muestra que es representativa de toda población; es debido aclarar que la población es el conglomerado de los individuos de interés para un estudio en particular y la muestra es el un subconjunto de la población. Para conocer y calcular el tamaño de la muestra se invita a revisar el documento **Tamaño de la muestra y cómo calcularlo.** | |

**Herramientas de tabulación**

| Cuadro de texto |
| --- |
| La tabulación es un proceso estadístico que permite agrupar los datos recolectados y presentarlos de una forma gráfica, así se facilita el análisis. Existen diversas herramientas que permiten realizar dicha acción, entre las que se destacan: |

| **Tipo de recurso** | Tarjetas Conectadas |
| --- | --- |
| **Introducción** |  |
| **Imagen:** 135401\_i40 | |
| **Ícono alusivo o representativo de la información**    **Imagen:** 135401\_i41 | **Excel**  Es una herramienta básica para el desarrollo de este tipo de actividades, permite calcular variables, realizar tablas y hacer gráficos. |
| **Imagen:** 135401\_i42 | ***Power BI***  Es una herramienta que permite importar datos y visualizarlos de manera dinámica, se caracteriza por ser un *software* amigable e interactivo y contribuye al análisis de datos. Esta herramienta incluye botones de comando que facilitan el proceso de visualización y segmentación de datos. |
| **Imagen:** 135401\_i43 | **SPSS**  Es un aplicativo de análisis estadístico que permite organizar y depurar los datos. Sin embargo, no es una herramienta intuitiva ni amigable, lo que genera que el uso sea más complejo. |
| **Imagen:** 135401\_i44 | ***Mintab***  Permite realizar pruebas y análisis estadísticos, es de fácil manejo y compatible con Excel ya que permite copiar datos directamente de dicho aplicativo. |

| Tipo de recurso | Cajón de texto de color |
| --- | --- |
| Los aplicativos mencionados permiten tabular y visualizar la información estadística**,** facilitando su sistematización y análisis; además automatizan los procesos, y se pueden articular con los sistemas de información organizacional contribuyendo al flujo de información. La elección del tipo de aplicativo depende de diversos criterios y dentro de los cuales se encuentran los requerimientos de manejo de la información a nivel de la organización, tecnología básica disponible, capacitación de los colaboradores y disponibilidad presupuestal. | |

| **Tipo de recurso** | Síntesis |
| --- | --- |
| Manejo de la información para sistemas de gestión  Síntesis: Recolección de información y bases de datos | |
| **Introducción** | A continuación, se presentan los principales contenidos del componente formativo, en él encontrará información relevante para iniciar la exploración de la estadística como herramienta que facilita los procesos de gestión y análisis de información. |
|  | |

| **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, material complementario, entre otros. Adicional, lo invitamos a resolver la actividad didáctica para reforzar los conceptos estudiados.** |

**MATERIAL COMPLEMENTARIO**

| **Tipo de recurso** | **Material complementario** | | |
| --- | --- | --- | --- |
| **Tema** | **Referencia APA del material** | **tipo** | **Enlace** |
| Tipos de Muestreo | Habiaspensado (2020). *¿Qué es el MUESTREO y TIPOS DE MUESTREO?* [Video]. Youtube. <https://www.youtube.com/watch?v=3LFDVSuaOkw> | Video | <https://www.youtube.com/watch?v=3LFDVSuaOkw> |
| Medidas de Dispersión | Matemóvil (2019). *Medidas de Dispersión o Variabilidad – Introducción.* [Video]. Youtube. <https://www.youtube.com/watch?v=Efg6G8vlVUA> | Video | <https://www.youtube.com/watch?v=Efg6G8vlVUA> |
| Tablas de Frecuencia | Daniel Carreón (2021). *Tabla de Frecuencias*. [Video]. Youtube. <https://www.youtube.com/watch?v=JtB2w0QLRZ4> | Video | <https://www.youtube.com/watch?v=JtB2w0QLRZ4> |
| Flujos de trabajo | GDX Group (2020). *Que es un Flujo de Trabajo y porque es importante en un Gestor Documental* (Video). Youtube<https://www.youtube.com/watch?v=Bph3b1t7ByU> | Video | <https://www.youtube.com/watch?v=Bph3b1t7ByU> |

**GLOSARIO**

| **Tipo de recurso** | Glosario |
| --- | --- |
| Análisis descriptivo: | proceso que permite hacer una descripción respecto a un conjunto de datos a partir de su comportamiento y características. |
| Base de datos: | herramienta que permite almacenar gran cantidad de datos, relacionados y estructurados, y pueden ser consultados rápidamente de acuerdo con las características selectivas que se deseen. |
| Datos: | información concreta sobre hechos, elementos, etc., que permiten ser estudiados, analizados o conocidos. |
| Documento: | es un elemento material o digital que alberga información importante para las organizaciones en el proceso de toma de decisiones. |
| Encuestas: | formulario estructurado que permite indagar a personas de interés de la organización sobre temas puntuales. |
| Estadística: | ciencia que utiliza conjuntos de datos numéricos para obtener, a partir de ellos, inferencias basadas en el cálculo de probabilidades. |
| Matriz: | cuadro que permite ordenar, almacenar y visualizar datos para ser graficados y analizados. |
| Observación: | técnica de recolección de datos que permite reconocer el comportamiento del objeto de estudio de manera directa. |
| Procesamiento: | manipular un conjunto de datos a organización, depuración y análisis ejecutando herramientas estadísticas. |
| Tabulación: | expresión de valores, magnitudes, conceptos, etc., por medio de tablas o cuadros. |

**ACTIVIDAD DIDÁCTICA**

| **Tipo de recurso** | Actividad didáctica. Arrastrar y soltar | |
| --- | --- | --- |
| En este ejercicio se evaluará la aprehensión de los conceptos desarrollados. Arrastre el concepto hasta la casilla con el enunciado o definición que corresponda: | | **Imagen:** 135401\_i45 |
| Disciplina de orden matemático que se centra en la interpretación de datos cuantitativos, permite analizar fenómenos reales o experimentales a través de la recolección y ordenamiento de una serie de datos. | | Estadística |
| Conjunto de procedimientos para la selección de una población determinada, y un subconjunto que se conoce como muestra; así generar conocimiento sobre las características de la población a través de la observación y generalización de esas características provenientes de la muestra. | | Muestreo |
| Recursos visuales que permiten comunicar la información. Se compone de un instrumento estadístico para analizar los datos en forma eficiente y desde diferentes niveles de lectura | | Gráficos estadísticos |
| Permite representar los elementos de la variable de estudio con los valores obtenidos en la frecuencia absoluta o relativa. En el eje X están las variables y en el eje Y el valor de las frecuencias. | | Gráfico de barras |
| Repositorio de información que se encuentra organizado de tal forma que permita al usuario acceder, gestionar y actualizar los datos de forma sencilla. | | Base de datos |
| Instrumento en el que se encuentran una serie de elementos que pueden ser objetos matemáticos y numéricos. | | Matriz |

**Retroalimentación positiva:**

¡Muy bien! Tiene claridad de los conceptos básicos estudiados en el componente formativo.

**Retroalimentación negativa:**

Parece que aún no tiene claros varios conceptos estudiados ¡Ánimo!. Se invita a repasar el componente formativo y realizar de nuevo la actividad interactiva.

**REFERENCIAS BIBLIOGRÁFICAS:**

| **Tipo de recurso** | Bibliografía |
| --- | --- |
| Bologna, E. (2018). *Métodos estadísticos de investigación*. Editorial Brujas. | |
| Campos, G. & Lule, N. (2012). *La observación, un método para el estudio de la realidad.* Revista Xihmai VII. | |
| Castellanos, M & Obando, J. (2021) *Gráficos estadísticos: guía práctica para estadística descriptiva*. Ediciones Universidad Cooperativa de Colombia. [2021\_GP\_Graficos estadisticos\_VF.pdf (ucc.edu.co)](https://repository.ucc.edu.co/bitstream/20.500.12494/36565/3/2021_GP_Graficos%20estadisticos_VF.pdf) | |
| Díaz, L., Torruco U., Martínez, M. & Varela, M.(2013). *Metodología de investigación en educación médica: La entrevista, recurso flexible y dinámico*. Investigación en Educación Médica. https://www.redalyc.org/pdf/3497/349733228009.pdf | |
| Dunstan, J. Maas, A. & Tobar, F. (2022). *Una mirada a la era de los datos*. Editorial Universitaria de Chile. | |
| Fernández, L.(2007) *¿Cómo se elabora un cuestionario?*. Butlletí LaRecerca. <https://www.ub.edu/idp/web/sites/default/files/fitxes/ficha8-cast.pdf> | |
| Gómez, I., Gualo, F., Merino, J., Muñoz-reja, I., Plattini M. & Rivas, B. (2018). Calidad de Datos. RA-MA. | |
| Guillermo, P. (s.f). *Noción y estructura del dato.* <http://metodos-comunicacion.sociales.uba.ar/wp-content/uploads/sites/219/2014/09/Abritta.pdf> | |
| Hamui, A. & Varela, M. (2013). *La técnica de grupos focales Investigación en Educación Médica*: *La técnica de grupos focales*. Investigación en Educación Médica. https://www.scielo.org.mx/pdf/iem/v2n5/v2n5a9.pdf | |
| Harari, Y. (2016). *Homo Deus. Breve historia del mañana.* Debate. <http://www.pratec.org/wpress/pdfs-pratec/Homo-Deus.pdf> | |
| Hernández, Z. (2012). *Métodos de análisis de datos: apuntes*. Universidad de La Rioja <https://dialnet.unirioja.es/servlet/libro?codigo=489791> | |
| Peña, D. (2015). *Fundamentos de estadística*. Alianza. | |
| Pulido, E., Escobar, Ó. & Núñez, J. (2019). *Base de datos*. Patria Educación. | |
| Ramos, F., & Guerra, R. (2019). *Introducción a los métodos estadísticos*. La Habana, Editorial Universitaria. | |
| Universidad Militar Nueva Granada (s.f). *Sistemas de Work Flow*. <http://virtual.umng.edu.co/distancia/ecosistema/ovas/ingenieria_industrial/sistemas_de_informacion_industrial/unidad_4/DM.pdf> | |