**FORMATO PARA EL DESARROLLO DE COMPONENTE FORMATIVO**

|  |  |
| --- | --- |
| PROGRAMA DE FORMACIÓN | Tecnología en Gestión Eficiente de la Energía |

|  |  |  |  |
| --- | --- | --- | --- |
| COMPETENCIA | 210601026 Procesar datos de acuerdo con procedimiento técnico y metodología estadística. | RESULTADOS DE APRENDIZAJE | 210601026-2. Procesar los datos recopilados empleando las herramientas informáticas que se ajustan a los requerimientos solicitados. |

|  |  |
| --- | --- |
| NÚMERO DEL COMPONENTE FORMATIVO | CF 18 |
| NOMBRE DEL COMPONENTE FORMATIVO | Procesamiento y análisis de datos |
| BREVE DESCRIPCIÓN | Un sistema de gestión eficiente es cuando los indicadores obtenidos son el resultado de un adecuado procesamiento y análisis de los datos. Es necesario desarrollar conocimiento y habilidades en diferentes métodos de procesamiento y análisis de datos para tener gran variedad al momento de determinar cuál es el adecuado dependiendo de los requerimientos del sistema. |
| PALABRAS CLAVE | Análisis, cualitativo, cuantitativo, datos, desviación. |

|  |  |
| --- | --- |
| ÁREA OCUPACIONAL | 2 - Ciencias naturales, aplicadas y relacionadas |
| IDIOMA | Español |

1. **TABLA DE CONTENIDOS:**

**Introducción**

**1. Probabilidad y estadística**

1.1. Moda, media, mediana y desviación típica, estudio de variables continuas

1.2. Distribuciones bidimensionales, diagramas de dispersión y rectas de regresión

1.3. Distribuciones discretas, distribución binomial, distribuciones continuas, distribución normal

1.4. Muestreo, distribución de medias muéstrales

1.5. Estimación y prueba de hipótesis

1.6. Formulario de muestreo y estimación

1.7. Probabilidad de sucesos compatibles e incompatibles

1.8. Cálculo de probabilidades y probabilidad condicionada

1.9. Combinatoria: Variaciones, permutaciones y combinaciones

**2. Métodos para procesar, graficar y analizar datos**

2.1. Métodos de investigación: métodos sintéticos, métodos analíticos, métodos inductivos, métodos deductivos, métodos comparativos

2.2. Métodos de procesamiento de datos más conocidos

2.3. Tipos de gráficas para el análisis de datos

2.4. Herramientas *software* más conocidas para procesar, graficar y analizar datos

**Introducción**

El procesamiento y el análisis de datos conllevan, anticipadamente, procedimientos de recopilación, limpieza y clasificación de estos, para luego llevar a cabo la transformación y el modelado de datos, bajo diferentes métodos tanto cualitativos como, cuantitativos con el objetivo de descubrir la información requerida.

Los resultados así obtenidos se comunican, sugiriendo conclusiones y apoyando la toma de decisiones. La visualización de datos se utiliza, a veces, para representar los datos para facilitar el descubrimiento de patrones útiles en los datos. Los términos Modelado de datos y Análisis de datos significan lo mismo.

Video motion

DI\_CF018\_0\_Introducción.

**1. Probabilidad y estadística**

La estadística forma parte del comportamiento histórico de todos los fenómenos naturales que ocurren en el mundo:

Cartas de diálogo

DI\_CF018\_1 Probabilidad y estadística.

* 1. **Moda, media, mediana y desviación típica, estudio de variables continuas**

A continuación, se definirá la moda, media, mediana y desviación típica, así como el estudio de variables continuas.

**Mediana**

La mediana es la observación intermedia en un conjunto de datos. Para contextualizar qué significa, se calcula la mediana de un conjunto de datos de muestra sobre el peso infantil, así:

Slyders

DI\_CF018\_1.1 Mediana.

**Moda**

Es el valor que aparece con mayor frecuencia en un conjunto de valores de datos. Si X es una variable aleatoria discreta, la moda es el valor x (es decir, X = x) en el que la función de masa de probabilidad toma su valor máximo. En otras palabras, es el valor que es más probable que se muestree. De esta manera:

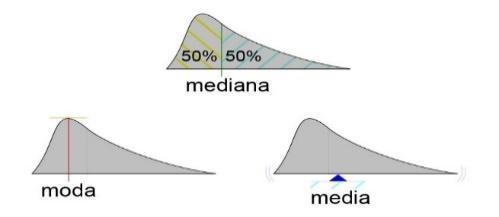
Presentación interactiva

DI\_CF018\_1.1\_Moda.

**Comparación de media, mediana y moda**

**Figura 1**

*Diferencias entre mediana, media y moda*



Nota. Tomada de <https://www.nobbot.com/educacion/no-te-quedes-en-la-media-y-aprende-a-calcular-la-mediana-asi-se-hace/>

A diferencia de la media y la mediana, el concepto de moda también tiene sentido para los “datos nominales” (es decir, que no constan de valores numéricos en el caso de la media, ni siquiera de valores ordenados en el caso de la mediana). Por ejemplo, tomando una muestra de apellidos coreanos, uno podría encontrar que “Kim” aparece con más frecuencia que cualquier otro nombre. Entonces “Kim” sería la moda de la muestra. En cualquier sistema de votación donde una pluralidad determina la victoria, un único valor modal determinar al vencedor, mientras que un resultado multimodal requeriría algún procedimiento de desempate.

A diferencia de la mediana, el concepto de moda tiene sentido para cualquier variable aleatoria que asuma valores de un espacio vectorial, incluidos los números reales (un espacio vectorial unidimensional) y los enteros (que pueden considerarse incrustados en los reales). Por ejemplo, una distribución de puntos en el plano normalmente tendrá una media y una moda, pero el concepto de mediana no se aplica. La mediana tiene sentido cuando existe un orden lineal en los valores posibles.

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Las generalizaciones del concepto de mediana a espacios de dimensiones superiores son la mediana geométrica y el punto central.  Tabla 1  *Descripción de la media aritmética, la mediana y la moda*   |  |  |  |  | | --- | --- | --- | --- | | Tipo | Descripción | Ejemplo | Resultado | | Media Aritmética | Suma de valores de un conjunto de datos dividida por el número de elementos o cantidad de números. | (1 + 2 + 2 + 3 + 4 + 7 + 9) / 7 | 4 | | Mediana | Valor medio o central que separa las mitades menor y mayor de un conjunto de elementos numéricos | 1, 2, 2, 3 , 4, 7, 9 | 3 | | Moda | Valor más frecuente en un conjunto de elementos numéricos. | 1, 2 , 2 , 3, 4, 7, 9 | 2 | |

**Varianza y desviación típica o estándar**

La varianza es la expectativa de la desviación al cuadrado de una variable aleatoria de su media poblacional o muestral. La varianza es una medida de dispersión, lo que significa que es una medida de qué tan lejos se separa un conjunto de números de su valor promedio y se define de manera detallada así:

Cartas de diálogo

DI\_CF018\_1.1\_\_Varianza y desviación típica o estándar.

La desviación estándar se usa ampliamente en entornos experimentales e industriales para probar modelos con datos del mundo real; un ejemplo de esto es el control de calidad de algunos productos. A continuación, se informa acerca de sus usos:

Presentación interactiva

DI\_CF018\_1.1\_\_\_ Varianza y desviación típica o estándar.

**Estudio de variables continuas**

Se invita a realizar un recorrido por el estudio de las variables continuas, así:

Slyders

DI\_CF018\_1.1\_\_\_\_Estudio de variables continuas

**1.2. Distribuciones bidimensionales, diagramas de dispersión y rectas de regresión**

En el siguiente recurso, se invita a descubrir todo sobre las distribuciones bidimensionales, diagramas de dispersión y rectas de regresión.

Slyders

DI\_CF018\_1.2 Distribuciones bidimensionales, diagramas de dispersión y rectas de regresión.

**Diagramas de dispersión**

Estas son herramientas matemáticas convenientes para estudiar la correlación entre dos variables aleatorias. Como su nombre indica, son una forma de hoja de papel sobre la que se encuentran dispersos los puntos de datos correspondientes a las variables de interés:

Slyders

DI\_CF018\_1.2\_Diagramas de dispersión.

**Rectas de regresión**

Están definidas de la siguiente manera:

Infografía dinámica

DI\_CF018\_1.2\_\_Rectas de regresión.

|  |
| --- |
| Para profundizar aún más en este ítem, diríjase al Anexo 1 - Distribuciones bidimensionales y rectas de regresión. |

* 1. **Distribuciones discretas, distribución binomial, distribuciones continuas, distribución normal**

A continuación, se exponen los diferentes tipos de distribución y cómo aplicarlos en la vida profesional.

Slyders

DI\_CF018\_1.3 Distribuciones discretas, distribución binomial, distribuciones continuas, distribución normal.

|  |
| --- |
| Distribución binomial  Suponer que X1, X2, …, Xn son variables aleatorias de Bernoulli independientes e idénticamente distribuidas (iid), cada una con la distribución    Entonces se dice que X tiene una distribución binomial con parámetros n y p:    Suponga que un experimento e x consiste en n ensayos repetidos tipo Bernoulli, cada ensayo resulta en un “éxito” con probabilidad π y un “fracaso” con probabilidad 1 – π. Por ejemplo, lanza una moneda 100 veces n = 100. Cuente el número de veces que observa cabezas, por ejemplo, X = # de cabezas. Si todos los ensayos son independientes, es decir, si la probabilidad de éxito en cualquier ensayo no se ve afectada por el resultado de cualquier otro ensayo, entonces el número total de éxitos en el experimento tendrá una distribución binomial, por ejemplo, dos lanzamientos de moneda no lo hacen afectarse unos a otros. La distribución binomial se puede escribir como    La distribución de Bernoulli es un caso especial del binomio con n = 1. Es decir,    Ello significa que X tiene una distribución de Bernoulli con probabilidad de éxito π. Se puede demostrar algebraicamente que si se cumple la ecuación anterior entonces:    Una forma más sencilla de llegar a estos resultados es tener en cuenta que dónde son variables aleatorias de Bernoulli. Entonces, por las propiedades aditivas de media y varianza, X= X1, X2, …, Xn donde X1, X2, …, Xn son variables aleatorias de Bernoulli. Entonces, por las propiedades aditivas de media y varianza,    y    Tenga en cuenta que X no tendrá una distribución binomial si la probabilidad de éxito π no es constante de un ensayo a otro, o si los ensayos no son completamente independientes (es decir, un éxito o fracaso en un ensayo altera la probabilidad de éxito en otro ensayo).    A medida que n aumenta, para fijo π, la distribución binomial se aproxima a la distribución normal:    Por ejemplo, si se toman muestras sin reemplazo de una población finita, entonces la distribución hipergeométrica es apropiada. |

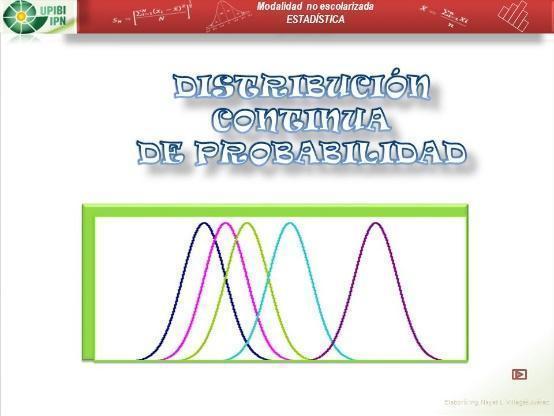
|  |
| --- |
| Para saber aún más sobre distribuciones discretas, diríjase al Anexo 2 - Modelos de probabilidad discretos. |

**Distribuciones continuas**

Estas pueden tomar cualquier valor dentro de un rango definido. Este rango puede ser infinito (por ejemplo, para la distribución normal) en cuyo caso se habla de una distribución limitada o si es finita (por ejemplo, la distribución uniforme) es una distribución delimitada.

**Figura 2**

*Distribuciones continuas*



Nota. Tomada de: <https://slideplayer.es/amp/1025784/>

|  |
| --- |
| En el Anexo 3 - Distribuciones continuas encontrarán una tabla que ofrece una descripción general de varias distribuciones continuas que se usan comúnmente en el modelado de análisis de riesgos, de modo que pueda enfocarse más fácilmente en cuáles podrían ser las más apropiadas para sus necesidades de modelado. Siga los enlaces para obtener una explicación detallada de cada uno. Se ha utilizado el nombre más común para cada distribución. |

Una distribución continua se utiliza para representar una variable que puede tomar cualquier valor dentro de un rango definido (dominio). Por ejemplo, la estatura de un inglés adulto escogido al azar tendrá una distribución continua, porque la estatura de una persona es esencialmente infinitamente divisible. Se podría medir su estatura en centímetro, milímetro, décimo de milímetro, etc.; la escala se puede dividir repetidamente generando cada vez más valores posibles.

Propiedades como el tiempo, la masa y la distancia, que son infinitamente divisibles, se modelan utilizando distribuciones continuas. En la práctica, también se usan distribuciones continuas para modelar variables que son, en verdad, discretas, pero donde la brecha entre los valores permitidos es insignificante: por ejemplo, el costo del proyecto (que es discreto con pasos de un centavo, un centavo, etc.), tipo de cambio (que solo se cotiza a unas pocas cifras significativas), número de empleados en una gran organización, etc.

|  |
| --- |
| Distribución normal  En la teoría de la probabilidad, la distribución normal (o gaussiana o Gauss o Laplace-Gauss) es un tipo de distribución de probabilidad continua para un valor real- variable aleatoria. La forma general de su función de densidad de probabilidad es:    El parámetro μ es la media o expectativa de la distribución (y también su mediana y moda), mientras que el parámetro es su desviación estándar σ. La varianza de la distribución es σ2. Se dice que una variable aleatoria con una distribución gaussiana tiene una distribución normal y se denomina desviación normal.  Figura 3  *Distribución normal*  https://4.bp.blogspot.com/-yIMsLCL_h30/V9b2PWVjm7I/AAAAAAAACuo/7XI4npuaYTMIymOoma7Oh_3qwvCZ5I2kQCLcB/s400/x.jpg  Nota. Tomada de <http://matepedia-estadistica.blogspot.com/2016/09/caracteristicas-de-una-distribucion.html>  Cualquier combinación lineal de una colección fija de desviaciones normales es una desviación normal. Muchos resultados y métodos, como la propagación de la incertidumbre y el ajuste de parámetros por mínimos cuadrados, pueden derivarse analíticamente de forma explícita cuando las variables relevantes se distribuyen normalmente. Una distribución normal a veces se denomina informalmente curva de campana. Sin embargo, muchas otras distribuciones tienen forma de campana (como las distribuciones de Cauchy, de Student y logística). |

|  |
| --- |
| Para profundizar aún más sobre distribuciones continuas, diríjase al Anexo 3 - Distribuciones continuas. |

**1.4. Muestreo, distribución de medias muéstrales**

Para este tema se invita a revisar de qué trata el muestreo y la distribución de medias muestrales.

Slyders

DI\_CF018\_1.4 Muestreo, distribución de medias muéstrales.

Y ahora se sugiere complementar la información con la siguiente presentación.

Presentación interactiva

DI\_CF018\_1.4\_Muestreo, distribución de medias muéstrales.

|  |
| --- |
| Distribución de medias muestrales  La media muestral  de un grupo de observaciones es una estimación de la media poblacional μ. Dada una muestra de tamaño n, considere n variables aleatorias independientes X1, X2, ..., Xn, cada una de las cuales corresponde a una observación seleccionada al azar. Cada una de estas variables tiene la distribución de la población, con media μ y desviación estándar σ. La media muestral se define como:  http://www.stat.yale.edu/Courses/1997-98/101/xbardef.gif  Por las propiedades de las medias y las varianzas de las variables aleatorias, la media y la varianza de la media muestral son las siguientes:  http://www.stat.yale.edu/Courses/1997-98/101/smmnvar.gif  Aunque la media de la distribución de  es idéntica a la media de la distribución de la población, la varianza es mucho menor para tamaños de muestra grandes.  Por ejemplo, suponga que la variable aleatoria X registra el puntaje de un estudiante seleccionado al azar en una prueba nacional, donde la distribución de la población para el puntaje es normal con una media de 70 y una desviación estándar de 5 (N (70,5)). Dada una muestra aleatoria simple (SRS) de 200 estudiantes, la distribución de la puntuación media de la muestra  tiene una media de 70 y una desviación estándar de 5 / sqrt (200) = 5 / 14,14 = 0,35.  La distribución de medias muestrales se define como el conjunto de medias de todas las posibles muestras aleatorias de un tamaño específico (n) seleccionadas de una población específica. Esta distribución tiene características bien definidas (y predecibles) que se especifican en el teorema del Límite central.  Figura 4  *Distribución de medias muestrales*    Nota. Tomada de <http://virtual.umng.edu.co/distancia/ecosistema/odin/odin_desktop.php?path=Li4vb3Zhcy9pbmdlbmllcmlhX2luZHVzdHJpYWwvZXN0YWRpc3RpY2FfaWkvdW5pZGFkXzEv#slide_2>  Cuando la distribución de la población es normal, la distribución de la media muestral también es normal. Para una distribución de población normal con media μ y desviación estándar σ, la distribución de la media de la muestra es normal, con media μ y desviación estándar.    Este resultado se deriva del hecho de que cualquier combinación lineal de variables aleatorias normales independientes también se distribuye normalmente. Esto significa que para dos variables aleatorias normales independientes X, Y y cualquier constante a y b, Ax + By serán distribuidos normalmente. En el caso de la media muestral, la combinación lineal es = (1 / n) \* (X1 + X2 + ... Xn). |

|  |
| --- |
| Teorema del Límite central  El resultado más importante sobre las medias muestrales es el teorema del Límite central. En pocas palabras, este teorema dice que para un tamaño de muestra n suficientemente grande, la distribución de la media muestral X se acercará a una distribución normal. Esto es cierto para una muestra de variables aleatorias independientes de cualquier distribución de población, siempre que la población tenga una desviación estándar finita σ. Una declaración formal del teorema del límite central es la siguiente:  Si X es la media de una muestra aleatoria X1, X2, ..., Xn de tamaño n de una distribución con una media finita μ y una varianza positiva finita σ², entonces la distribución de W =  http://www.stat.yale.edu/Courses/1997-98/101/clt.gif  es N (0,1) en el límite cuando n se acerca al infinito.  Esto significa que la variable  se distribuye N (μ, http://www.stat.yale.edu/Courses/1997-98/101/smvar.gif).  Una aplicación bien conocida de este teorema es la aproximación normal a la distribución binomial.  Figura 5  *Teorema del Límite central*    Nota. Tomada de <http://virtual.umng.edu.co/distancia/ecosistema/odin/odin_desktop.php?path=Li4vb3Zhcy9pbmdlbmllcmlhX2luZHVzdHJpYWwvZXN0YWRpc3RpY2FfaWkvdW5pZGFkXzEv#slide_3> |

El concepto de distribución de medias muestrales y sus características deben ser intuitivamente razonables:

* Es de notar que las medias muestrales son variables. Si se seleccionan dos (o más) muestras de la misma población, es probable que las dos muestras tienen medias diferentes.
* Aunque las muestras tendrán medias diferentes, se debe esperar que las medias muestrales estén cerca de la media de la población. Es decir, las medias de la muestra deberían “acumularse” alrededor de μ. Por tanto, la distribución de las medias muestrales tiende a adoptar una forma normal con un valor esperado de μ.
* Debe darse cuenta de que la media de una muestra individual probablemente no será idéntica a la media de su población; es decir, habrá algún “error” entre X y μ. Algunas medias de muestra estarán relativamente cerca de μ y otras relativamente lejos. El error estándar proporciona una medida de la distancia estándar entre X y μ.

|  |
| --- |
| Puntuaciones Z y ubicación dentro de la distribución de medias muestrales   * Dentro de la distribución de las medias muestrales, la ubicación de cada media muestral se puede especificar mediante una puntuación Z,   https://sphweb.bumc.bu.edu/otlt/mph-modules/bs/sas/sas4-onesamplettest/Equation8.png   * Como siempre, una puntuación Z positiva indica una media muestral mayor que μ y una puntuación Z negativa corresponde a una media muestral menor que μ.   El valor numérico de la puntuación Z indica la distancia entre  y μ medida en términos del error estándar. |

**1.5. Estimación y prueba de hipótesis**

Un área de preocupación en la estadística inferencial es la estimación del parámetro de población a partir de la estadística de la muestra; aquí es importante darse cuenta del orden. La estadística de muestra se calcula a partir de los datos de muestra y el parámetro de población se infiere (o estima) a partir de esta estadística de muestra:

Slyders

DI\_CF018\_1.5 Estimación y prueba de hipótesis.

Cinco pasos en la prueba de hipótesis:

* Especificar la hipótesis nula.
* Especificar la hipótesis alternativa.
* Establecer el nivel de significación (α).
* Calcule la estadística de prueba y el valor P correspondiente.
* Esbozando una conclusión.

Para ampliar la información sobre estos importantes pasos en la estimación y la prueba de hipótesis, se invita a revisar:

Pasos

DI\_CF018\_1.5\_Estimación y prueba de hipótesis.

**Precauciones sobre los valores p**

El tamaño de la muestra afecta directamente su valor p. Los tamaños de muestra grandes producen valores p pequeños incluso cuando las diferencias entre los grupos no son significativas. Siempre debe verificar la relevancia práctica de sus resultados. Por otro lado, un tamaño de muestra demasiado pequeño puede provocar que no se identifique una diferencia cuando realmente existe.

Se debe planificar el tamaño de la muestra con anticipación para tener suficiente información de su muestra para mostrar una relación o diferencia significativa, si existe. Consulte el cálculo de un tamaño de muestra para obtener más información.

**Ejemplo**

* Las edades medias fueron significativamente diferentes entre los dos grupos (16,2 años frente a 16,7 años; p = 0,01; n = 1.000). ¿Es esta una diferencia importante? Probablemente no, pero el gran tamaño de la muestra ha dado como resultado un valor p pequeño.

**Ejemplo**

* Las edades medias no fueron significativamente diferentes entre los dos grupos (10,4 años frente a 16,7 años; p = 0,40, n = 10). ¿Es esta una diferencia importante? Podría ser, pero debido a que el tamaño de la muestra es pequeño, no se puede determinar con certeza si se trata de una verdadera diferencia o simplemente sucedió debido a la variabilidad natural en la edad dentro de estos dos grupos.

Si realiza una gran cantidad de pruebas para evaluar una hipótesis (lo que se denomina prueba múltiple), debe controlar esto en su designación del nivel de significancia o en el cálculo del valor p. Por ejemplo, si tres resultados miden la efectividad de un fármaco u otra intervención, tendrá que ajustar estos tres análisis.

|  |
| --- |
| Sacar una conclusión  1. Valor p < = nivel de significancia (a) = >, rechaza su hipótesis nula a favor de su hipótesis alternativa. El resultado es estadísticamente significativo.  2. Valor p > nivel de significancia (a) = > no rechaza su hipótesis nula. El resultado no es estadísticamente significativo.  La prueba de hipótesis no está configurada para que pueda probar absolutamente una hipótesis nula. Por lo tanto, cuando no se encuentra evidencia en contra de la hipótesis nula, no se puede rechazar la hipótesis nula. Cuando se encuentra evidencia suficientemente fuerte en contra de la hipótesis nula, se rechaza la hipótesis nula. Sus conclusiones también se traducen en una declaración sobre su hipótesis alternativa. Cuando presente los resultados de una prueba de hipótesis, incluya también las estadísticas descriptivas en sus conclusiones. Informe valores p exactos en lugar de un rango determinado. Por ejemplo, “La tasa de intubación difirió significativamente según la edad del paciente y los pacientes más jóvenes tienen una tasa más baja de intubación exitosa (p = 0,02)”. Aquí hay dos ejemplos más con la conclusión expresada de varias formas diferentes.  Ejemplo   * H0: no hay diferencia en la supervivencia entre el grupo de intervención y el de control. * HA: existe una diferencia en la supervivencia entre el grupo de intervención y el de control. * α = 0,05: aumento del 20% en la supervivencia para el grupo de intervención; valor p = 0,002.   Conclusión   * Rechace la hipótesis nula a favor de la hipótesis alternativa. * La diferencia en la supervivencia entre el grupo de intervención y el de control fue estadísticamente significativa. * Hubo un aumento del 20% en la supervivencia para el grupo de intervención en comparación con el control (p = 0,001).   Ejemplo   * H0: no hay diferencia en la supervivencia entre el grupo de intervención y el de control. * HA: existe una diferencia en la supervivencia entre el grupo de intervención y el de control. * α = 0,05: aumento del 5% en la supervivencia entre el grupo de intervención y el de control; valor p = 0,20.   Conclusión   * No rechace la hipótesis nula. * La diferencia en la supervivencia entre el grupo de intervención y el de control no fue estadísticamente significativa. * No hubo un aumento significativo en la supervivencia para el grupo de intervención en comparación con el control (p = 0,20). |

**1.6. Formulario de muestreo y estimación**

Las fórmulas que se usan para calcular la media muestral y todas las demás estadísticas muestrales son ejemplos de fórmulas de estimación o estimadores. El valor particular que calculamos a partir de observaciones de muestra utilizando un estimador se denomina estimación. Por ejemplo, el valor calculado de la media muestral en una muestra determinada se denomina estimación puntual de la media poblacional. Las tres propiedades deseables de un estimador son:

* **Insesgado:** su valor esperado es igual al parámetro que se está estimando.
* **Eficiencia:** tiene la varianza más baja en comparación con otros estimadores insesgados del mismo parámetro.
* **Coherencia:** a medida que aumenta el tamaño de la muestra, el error muestral disminuye y las estimaciones se acercan al valor real.

|  |
| --- |
| En el Anexo 4 - Formulario de muestreo, se puede encontrar una clasificación por tablas de fórmulas más utilizadas y necesarias para llevar a cabo todo lo relacionado a procesos de muestreo de datos. |

**1.7. Probabilidad de sucesos compatibles e incompatibles**

Dos o más eventos son compatibles, si pueden cumplirse simultáneamente; es decir, si tienen al menos un resultado común. Por el contrario, son incompatibles o mutuamente excluyentes y su intersección es el conjunto vacío Ø si se analiza el siguiente experimento de lanzar un dado.

|  |
| --- |
| Comencemos con el siguiente experimento: tiramos un dado de seis caras y vemos cuál es el resultado. Consideremos los siguientes eventos A= {2, 3}, B= {1,2}, C= {5}.  Observamos que, si extraemos 2, luego A está satisfecho, así como B. Decimos que los eventos son compatibles, esto significa que pueden ocurrir simultáneamente. Por el contrario, los eventos B y C son incompatibles, ya que los dos no pueden suceder simultáneamente.  Para ver cuando dos eventos son compatibles o no, podemos observar que A y B tienen un elemento común: 2, por lo tanto, serán compatibles. De lo contrario, A y C no tienen ningún elemento común y, por tanto, son incompatibles. Expresamos esto diciendo que dos eventos A y B son compatibles si:  A Ո B ≠ Ø  y, por el contrario, son incompatibles si:  A Ո B = Ø  Si tenemos tres o más eventos, decimos que son incompatibles de dos en dos si dos eventos son incompatibles (de manera similar, son compatibles de dos en dos si dos eventos son compatibles). En nuestro caso A, B y C no son incompatibles de dos en dos, ya que, aunque A y C, al igual que B y C son incompatibles, A y B son compatibles. |

Ahora, se invita al siguiente recurso:

Tarjetas

DI\_CF018\_1.7 Probabilidad de sucesos compatibles e incompatibles.

**1.8. Cálculo de probabilidades y probabilidad condicionada.**

Una probabilidad es un modelo representado matemáticamente de un fenómeno aleatorio. Las probabilidades pueden ser marginales, conjuntas o condicionales.

Slyders

DI\_CF018\_1.8 Cálculo de probabilidades y probabilidad condicionada.

|  |
| --- |
| Cómo calcular la probabilidad de ocurrencia de A o B, si los eventos son incompatibles: Si A y B son dos eventos incompatibles, es decir, no pueden ocurrir al mismo tiempo, la probabilidad de que ocurra A o B será la suma de las probabilidades de que cada evento ocurra por separado.  Cómo calcular la probabilidad  La probabilidad de obtener una bola blanca es:  calcular probabilidad  La probabilidad de obtener una bola negra es:  cálculo de probabilidad  Por lo tanto, la probabilidad de obtener una bola blanca o una bola negra es:  ley de laplace probabilidad  ley de laplace |

|  |
| --- |
| Cómo calcular la probabilidad de que ocurra A o B, si los eventos son compatibles: si A y B son dos eventos compatibles, es decir, pueden ocurrir al mismo tiempo, entonces la probabilidad de que ocurran A o B será:  suma de probabilidades  P (A ∩ B) dice: la probabilidad de que ocurran A y B.  Esta vez, la suma de las probabilidades de que cada evento ocurra por separado debe restarse de la probabilidad de que los dos eventos ocurran al mismo tiempo. Por ejemplo, calcular la probabilidad de que al lanzar un dado el número obtenido sea par o que sea un 4. En este caso, el evento “obtener un número par” y el evento “obtener un 4” son compatibles, porque si obtenemos un 4 están sucediendo ambos eventos a la vez. Por lo tanto, la probabilidad de obtener 4 o un número par se calculará con la fórmula:  eventos incompatibles  La probabilidad de obtener un número par es:  probabilidad de eventos compatibles  La probabilidad de obtener un 4 es:  probabilidad compatible e incompatible  La probabilidad de obtener un número par y un 4 es:  ley de probabilidades  Solo hay una posibilidad entre 6 ya que 4 es el único número que cumple ambos eventos al mismo tiempo. Finalmente, la probabilidad de obtener un número par o 4 es:  aprender las probabilidades paso a paso  Que la sustitución de cada término por su valor nos queda a nosotros:  tipos de eventos probabilidad |

**Probabilidad condicionada**

En la teoría de la probabilidad, la probabilidad condicionada o condicional es una medida de la probabilidad de que ocurra un evento, dado que ya ha ocurrido otro evento (por suposición, presunción, afirmación o evidencia). Veamos:

Presentación interactiva

DI\_CF018\_1.8\_Probabilidad condicionada.

**1.9. Combinatoria: variaciones, permutaciones y combinaciones**

La combinatoria se ocupa de conjuntos finitos y sus objetos básicos son permutaciones, variaciones y combinaciones (y se refieren tanto a las operaciones como a los resultados de estas operaciones). Dado que siempre es posible identificar m y n elementos conjuntos con los conjuntos {1, 2, ..., m} y {1, 2, ..., n}, es suficiente considerar estos conjuntos, y esta situación es una a uno estándar. Recordamos estas tres nociones (en sus dos apariencias: cuándo se permiten las repeticiones y cuándo no).

Esto se hace comenzando con variaciones sin repeticiones (validando esta decisión por una relación directa con la cardinalidad de todos los mapas finitos), aunque se podría comenzar con combinaciones o con permutaciones. Al cumplir con la permutación, hay que prestar una especial atención a las operaciones con factorial.

|  |
| --- |
| En el Anexo 5 - Combinatoria - Variaciones, permutaciones y combinaciones, se encuentran las respectivas fórmulas para un caso determinado y su respectivo ejemplo. |

**2. Métodos para procesar, graficar y analizar datos**

Después de la recolección de datos, limpieza inicial y clasificación de los mismos, se sigue un conjunto de procedimientos para obtener los datos / información deseados del trabajo de campo que se desea ejecutar, para procesar, graficar y analizar los hechos de manera lógica y científica.

**Figura 6**

*Procesamiento, análisis y gráfica de datos*



Nota. Tomada de <http://metodologiayvidaucv.blogspot.com/2019/02/procesamiento-de-datos.html>

**2.1. Métodos de investigación**

Estos métodos son procesos que se utilizan para recopilar datos. Puede utilizar estos datos para analizar métodos o procedimientos actuales y encontrar información adicional sobre un tema. Los profesionales utilizan métodos de investigación mientras estudian medicina, comportamiento humano y otros temas académicos. Hay dos categorías principales de métodos de investigación: métodos de investigación cualitativa y métodos de investigación cuantitativa; ahora se ejemplifica la mirada cuantitativa:

Cartas de diálogo

DI\_CF018\_2.1\_Métodos de investigación.

**2.2. Métodos de procesamiento de datos más conocidos**

**Figura 7**

*Procesamiento de datos*



Nota. Tomada de <https://www.nextu.com/blog/que-es-el-procesamiento-de-datos/>

La recopilación, manipulación y procesamiento de datos recopilados para el uso requerido se conoce como procesamiento de datos, es una técnica realizada normalmente por una computadora; el proceso incluye la recuperación, transformación o clasificación de información. Sin embargo, el procesamiento de datos depende en gran medida de lo siguiente:

* El volumen de datos que deben procesarse.
* La complejidad de las operaciones de procesamiento de datos.
* Capacidad y tecnología incorporada del sistema informático respectivo.
* Habilidades técnicas.
* Limitaciones de tiempo.

Para completar la temática, se invita a realizar un recorrido por los diferentes métodos de procesamiento de datos:

Presentación interactiva

DI\_CF018\_2.2 Métodos de procesamiento de datos más conocidos.

**2.3. Tipos de gráficas para el análisis de datos**

Es necesario conocer las diferentes gráficas para realizar un análisis de datos; a saber:

Tarjetas

DI\_CF018\_2.3\_Tipos de gráficas para el análisis de datos

Desde las métricas del sitio web y el desempeño del equipo de ventas hasta los resultados de las campañas de *marketing* y las tasas de adopción de productos, existe una variedad de puntos de datos que su organización necesita rastrear. Cuando se tienen las manos ocupadas haciendo malabares con varios proyectos a la vez, se necesita un método de informes rápido y eficaz que le permita expresar un punto claro. ¿Se sabe qué tipo de método de visualización de datos utilizar?

Algunos de los tipos más comunes de gráficos y gráficos de visualización de datos incluyen:

* Gráfico de columnas.
* Gráfico de barras.
* Gráfico de barras apiladas.
* Gráfico de columnas apiladas.
* Gráfico de área.
* Gráfico de doble eje.
* Gráfico de líneas.
* Gráfico de Mekko.
* Gráfico circular.
* Gráfico de cascada.
* Gráfico de burbujas.
* Gráfico de diagrama de dispersión.
* Gráfico de viñetas.
* Gráfico de embudo.
* Mapa de calor.

Si bien todos sirven para agilizar y mejorar la interpretación de datos, no todos son apropiados para el mismo trabajo. Elegir la ayuda visual adecuada es la clave para evitar la confusión del usuario y asegurarse de que su análisis sea preciso. Ahora se analizarán 10 de estos 15 tipos de tablas y gráficos.

**Explicación de 10 tipos de visualización de datos**

Existen innumerables tipos diferentes de tablas, gráficos y otras técnicas de visualización que pueden ayudar a los analistas a representar y transmitir datos importantes. Echemos un vistazo a 10 de los más comunes:

|  |
| --- |
| Gráfico de columnas  Este es uno de los tipos más comunes de herramientas de visualización de datos. Hay una razón por la que aprendemos a hacer gráficos de columnas en la escuela primaria. Son una forma simple y tradicional de mostrar una comparación entre diferentes conjuntos de datos. También se puede utilizar un gráfico de columnas para realizar un seguimiento de los conjuntos de datos a lo largo del tiempo.  Figura 8  *Gráfico de barras*  Gráfico de barras en Excel 2013  Nota. Tomada de <https://dtutivenc20162017.wordpress.com/grafico-de-barras-y-columnas/>  Un gráfico de columnas incluirá etiquetas de datos a lo largo del eje horizontal (X) con métricas o valores medidos presentados en el eje vertical (Y), también conocido como el lado izquierdo del gráfico. El eje Y normalmente comenzará en 0 y llegará tan alto como la medida más grande que está rastreando.  Puede utilizar gráficos de columnas para realizar un seguimiento de las cifras de ventas mensuales, los ingresos por página de destino o medidas similares. Los colores consistentes ayudan a mantener el enfoque en los datos en sí, aunque puede introducir colores de acento para enfatizar puntos de datos importantes o para realizar un seguimiento de los cambios a lo largo del tiempo. |

|  |
| --- |
| Gráfico de barras  A menudo puede usar un gráfico de barras y un gráfico de columnas de la misma manera, aunque los gráficos de columnas limitan su etiqueta y el espacio de comparación. Es mejor seguir con un gráfico de barras si:   * Trabaja con etiquetas más largas. * Visualiza números negativos. * Compara 10 o más elementos.   Figura 9  *Gráfico de barras*  Cómo realizar un gráfico de barras en Excel 2007  Nota. Tomada de <https://www.mundodeportivo.com/uncomo/tecnologia/articulo/como-realizar-un-grafico-de-barras-en-excel-2007-15090.html>  En este caso, sus etiquetas de datos irán a lo largo del eje Y, mientras que las medidas estarán a lo largo del eje X. |

|  |
| --- |
| Gráfico de barras apiladas  ¿Está comparando muchos elementos diferentes? ¿Desea realizar un seguimiento del crecimiento individual de cada conjunto de datos en sí, junto con el crecimiento del grupo como un todo colectivo? Para revelar esta relación de parte a todo, se creará un gráfico de barras apiladas.  Al eliminar el color de este gráfico, se vería similar a un gráfico de barras estándar. El diseño “apilado” representa la combinación de colores contrastantes de este gráfico. Estos colores se asignan a una leyenda que acompaña a su mapa.  Por ejemplo, es posible que desee realizar un seguimiento del rendimiento de cuatro tipos diferentes de productos en cinco estrategias de ventas diferentes. La estrategia 1 a la estrategia 5 estará en su eje X, mientras que las cifras de ventas estarán en el eje Y.  Figura 10  *Gráfico de barras apiladas*  Gráfico de barras apiladas según la ciudad y el TIV, subagrupado por clases de póliza  Nota. Tomada de <https://doc.arcgis.com/es/insights/latest/create/stacked-bar-chart.htm>  Sin embargo, dentro de cada categoría de estrategia, tendrá cuatro bloques de colores diferentes. Cada uno representa uno de los tipos de productos. De esta manera, puede determinar qué estrategia funcionó mejor para cada tipo de producto en su conjunto, así como qué productos funcionaron bien dentro de cada estrategia. |

|  |
| --- |
| Gráfico de líneas  Este un tipo de gráficos estándar que se reconoce al instante. Un gráfico de líneas está diseñado para revelar tendencias, avances o cambios que ocurren a lo largo del tiempo. Como tal, funciona mejor cuando el conjunto de datos es continuo en lugar de estar lleno de arranques y paradas.  Figura 11  *Gráfico de líneas*  http://www.gusgsm.com/files/graficolinea_06.png  Nota. Tomada de <http://www.gusgsm.com/graficos_de_linea_y_area_en_adobe_illustrator>  Como un gráfico de columnas, las etiquetas de datos en un gráfico de líneas están en el eje X mientras que las medidas están en el eje Y.  Se deben usar líneas continuas y evite trazar más de cuatro líneas, ya que cualquier cosa por encima de esto puede distraer; además, hay que planificar el espacio suficiente para que tus líneas midan alrededor de 2/3 de la altura del eje Y. |

|  |
| --- |
| Gráfico de doble eje  Este tipo de gráfico se utiliza para presentar la relación entre dos variables. Más específicamente, estos gráficos son útiles para demostrar la relación entre dos o más medidas con diferente amplitud y escala. Los gráficos de doble eje ayudan a presentar mucha información en el espacio limitado presente en su tablero y también le permiten comprender las tendencias que, de lo contrario, podría haber perdido.  Figura 12  *Gráfico de doble eje*  gráfico de doble eje  Nota. Tomada de <https://www.saberprogramas.com/excel-como-crear-graficos-de-doble-eje-en-excel/> |

|  |
| --- |
| Gráfico Mekko  Este es un gráfico con el que podría estar menos familiarizado, a menos que se encuentre en el espacio de análisis de datos. Un gráfico de Mekko tiene un diseño similar a un gráfico de barras apiladas, con una excepción importante: en lugar de seguir la progresión del tiempo, el eje X mide otra dimensión de sus conjuntos de datos.  Figura 13  *Gráfico Mekko*  Infografía de gráfico lineal mekko vector gratuito  Nota. Tomada de <https://www.freepik.es/vector-gratis/infografia-grafico-lineal-mekko_10371574.htm>  Con este diseño, puede comparar valores, medir la composición de cada valor y analizar la distribución de datos, todo al mismo tiempo. |

**Gráfico circular**

Un gráfico circular representa un número estático, dividido en categorías que constituyen sus porciones individuales; este representará cantidades numéricas en porcentajes. Cuando sume todas las porciones separadas, deberían sumar el 100%. Estos son especialmente útiles en *marketing* digital, ya que puede usarlos para mostrar un desglose de:

* Cuotas de mercado.
* Gastos de *marketing*.
* Demografía del cliente.
* Uso del dispositivo del cliente (para pruebas de UX).
* Fuentes de tráfico *online*.

Si se desea que su gráfico circular tenga mucha diferenciación entre los sectores, es mejor limitar la cantidad de categorías que ilustra.

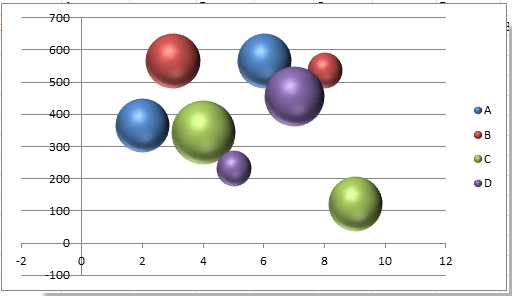
|  |
| --- |
| Diagrama de dispersión  Este tipo de visualización también se denomina diagrama de dispersión y representa diferentes variables trazadas a lo largo de dos ejes. Tenga en cuenta que tanto el eje X como el eje Y son ejes de valor, ya que un gráfico de dispersión no utiliza un eje de categoría.  Figura 14  *Diagrama de dispersión*    Nota. Tomada de <http://www.disfrutalasmatematicas.com/datos/grafica-dispersion-xy.html>  Estos tipos de visualización de datos funcionan mejor cuando analiza varios puntos de datos y busca similitudes dentro del conjunto de datos. Al hacerlo, se puede notar cualquier valor atípico y también obtener una comprensión más clara de la distribución general de sus datos.  Digamos, por ejemplo, que desea medir las puntuaciones de los comentarios de los clientes que recibe su organización. También quería ver si los tiempos de respuesta de su mesa de servicio tienen algún impacto en esos puntajes. Los puntajes de retroalimentación varían de 0 a 10, por lo que esas serían sus medidas del eje Y. En su eje X, etiquetaría desde 0 hasta el tiempo de respuesta más largo permitido, como una hora. Luego, trazaría los puntajes que recibió, notando patrones y tendencias que pueden ayudarlo a informar sus esfuerzos de servicio. |

**Gráfico de burbujas**

Al igual que un gráfico de dispersión, un gráfico de burbujas también puede mostrar relaciones o distribución. En esta variación, sin embargo, reemplazará los puntos de datos con burbujas. También variará los tamaños de la burbuja para representar un tercer conjunto de datos.

**Figura 15**

*Gráfico de burbujas*



Nota. Tomada de <https://es.extendoffice.com/documents/excel/2347-excel-bubble-chart-color-based-on-value.html>

Al igual que con un gráfico de dispersión, uno de burbujas no utiliza un eje de categorías. Más bien, trazará los conjuntos de datos como valores X, valores Y y ahora, valores Z (tamaño de burbuja).

|  |
| --- |
| Gráfico de viñetas  Este tipo de gráfico puede ayudar a realizar un seguimiento visual de su progreso. Similar al diseño gráfico de barras, estos también incorporan otros elementos visuales. Al usar un gráfico de viñetas, comenzará con una medida principal, y luego comparará esa medida con otra (o múltiple) para encontrar un significado y una conexión más profundos.  Figura 16  *Gráfico o de viñetas*  Bullet chart visualization.  Nota. Tomada de <https://help.qlik.com/es-ES/sense/August2021/Subsystems/Hub/Content/Sense_Hub/Visualizations/VisualizationBundle/bullet-chart.htm>  Ahora que hemos explorado los diferentes tipos de gráficos, cuadros y mapas de visualización de datos, analicemos brevemente algunas de las razones por las que podría necesitar la visualización de datos en primer lugar. Si no sabe qué tipo de visual funcionará mejor para su empresa, es útil comprender las principales funciones comerciales que puede ofrecer la visualización de datos. Aquí están los cinco principales a considerar: |

Acordeón

DI\_CF018\_2.3\_Cinco razones esenciales para implementar herramientas de visualización de datos.

**2.4. Herramientas *software* más conocidas para procesar, graficar y analizar datos**

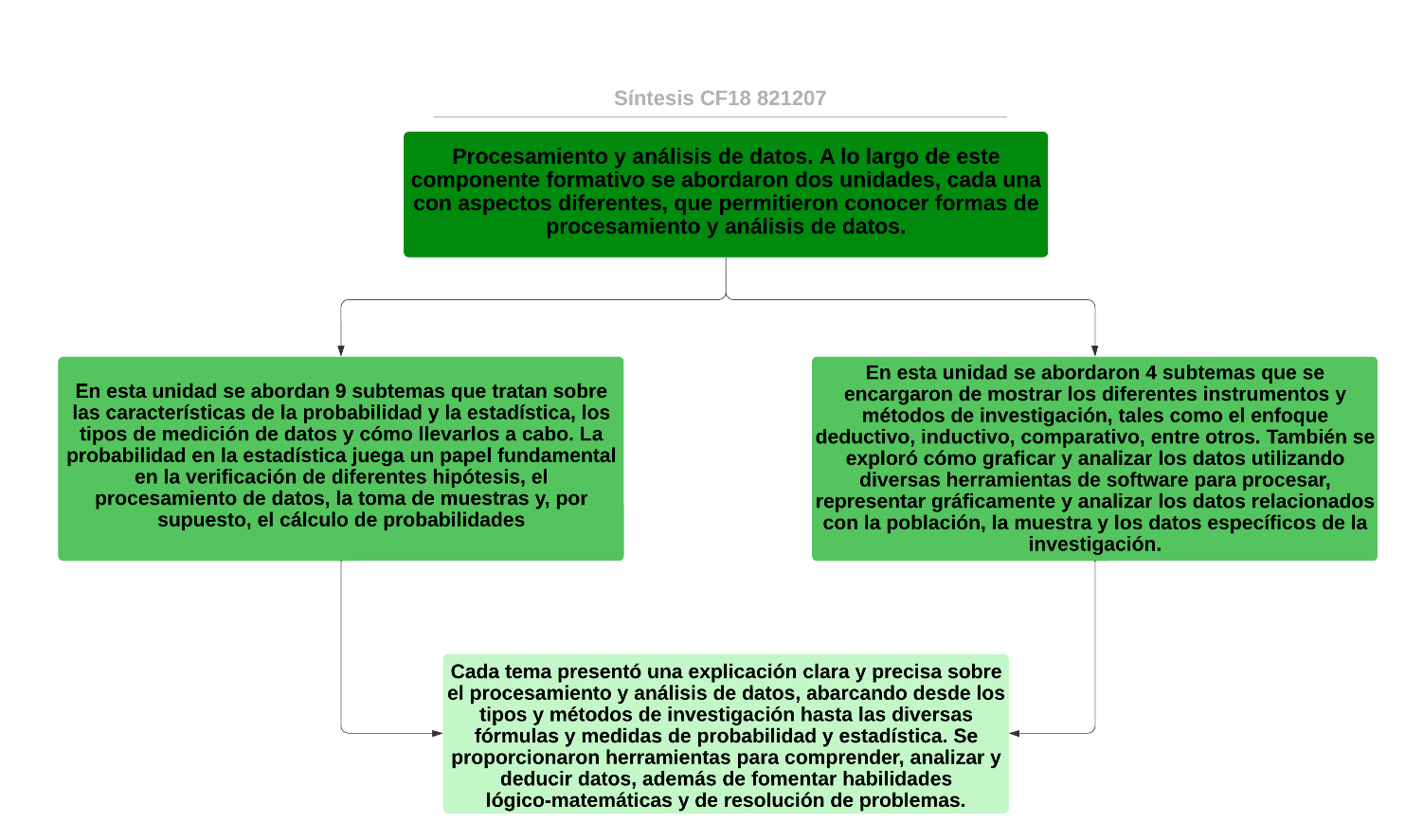
A continuación, encontrará las herramientas necesarias para procesar, graficar y analizar los datos:

Slyders

DI\_CF018\_2.4 Herramientas software más conocidas para procesar, graficar y analizar datos.

|  |
| --- |
| Para tener una idea más clara y profunda de esta temática, diríjase al Anexo 6 - Principales herramientas y *software* de análisis de datos 2021. |

**Síntesis**

El procesamiento y análisis de datos son elementos clave en la toma de decisiones basada en información cuantitativa. Dentro de este campo, se abordan diversos subtemas, como la probabilidad y estadística, que incluyen el estudio de variables continuas, distribuciones bidimensionales, discretas y continuas, así como muestreo, estimación, prueba de hipótesis y combinatoria. Además, se exploran métodos para procesar, graficar y analizar datos, como métodos de investigación, tipos de gráficas y herramientas software ampliamente utilizadas en esta área. Estas habilidades permiten comprender mejor los datos, identificar patrones y tomar decisiones informadas en diversos campos.

**ACTIVIDADES DIDÁCTICAS (OPCIONALES SI SON SUGERIDAS)**

|  |  |
| --- | --- |
| DESCRIPCIÓN DE ACTIVIDAD DIDÁCTICA | |
| Nombre de la Actividad | Explorando Variables: Continuas vs. Discretas |
| Objetivo de la actividad | El objetivo de esta actividad didáctica es que los aprendices puedan comprender y distinguir entre variables continuas y variables discretas, identificando ejemplos apropiados de cada una. |
| Tipo de actividad sugerida | Cuestionario |
| Archivo de la actividad  (Anexo donde se describe la actividad propuesta) | CF18-Actividad didáctica cuestionario CF18 |

1. **MATERIAL COMPLEMENTARIO:**

|  |  |  |  |
| --- | --- | --- | --- |
| Tema | Referencia APA del Material | Tipo de material  (Video, capítulo de libro, artículo, otro) | Enlace del Recurso o  Archivo del documento o material |
| 2.2. Métodos de procesamiento de datos más conocidos. | Tesis de Cero a 100 – TV. (2019). *Guía básica para análisis estadístico de datos | Parte 1*. [Video]. YouTube. https://www.youtube.com/watch?v=Wn9gQjWNdiY | Video | <https://www.youtube.com/watch?v=Wn9gQjWNdiY> |
| 2.2. Métodos de procesamiento de datos más conocidos. | Tesis de Cero a 100 – TV (2019). *Guía básica análisis estadístico de datos | Parte 2*. [Video]. YouTube. https://www.youtube.com/watch?v=CnxMorGipGw | Video | <https://www.youtube.com/watch?v=CnxMorGipGw> |
| 2.2. Métodos de procesamiento de datos más conocidos. | Comunicación numérica. (2020). *Fundamentos del análisis de datos para toma de decisiones*. [Video]. YouTube. https://www.youtube.com/watch?v=qvZxvMWMvDo | Video | <https://www.youtube.com/watch?v=qvZxvMWMvDo> |
| 2.3. Tipos de gráficas para el análisis de datos. | A2 Capacitación: Excel. (2019). *Minicurso de business intelligence en Excel - tablas dinámicas, gráficas y dashboards - Parte 1*. [Video]. YouTube. https://www.youtube.com/watch?v=TgcGa0cnIpg | Video | <https://www.youtube.com/watch?v=TgcGa0cnIpg> |
| 2.3. Tipos de gráficas para el análisis de datos. | A2 Capacitación: Excel. (2019). *Minicurso de business intelligence en Excel - tablas dinámicas, gráficas y dashboards -* - *Parte 2*. [Video]. YouTube. https://www.youtube.com/watch?v=ucJM-wpuoKg | Video | <https://www.youtube.com/watch?v=ucJM-wpuoKg> |
| Herramientas *software* más conocidas para procesar, graficar y analizar datos. | Las mates fáciles. (2020). *Probabilidad: sucesos compatibles e incompatibles – Explicación.* [Video]. YouTube. https://www.youtube.com/watch?v=njjrzLk5RB0 | Video | <https://www.youtube.com/watch?v=njjrzLk5RB0> |

1. **GLOSARIO:**

|  |  |
| --- | --- |
| TÉRMINO | SIGNIFICADO |
| Amplitud o rango | diferencia entre el valor máximo y mínimo de los valores de una variable se encuentran comprendidos el 100% de los valores muestrales. |
| Área bajo la curva (entre dos puntos) | si la curva viene dada por una función de densidad teórica, representa la probabilidad de que la variable aleatoria tome un valor dentro del intervalo determinado por esos dos puntos. |
| Características | propiedades de las unidades o elementos que componen las muestras. Se miden mediante variables. Se asume que los individuos presentan diferentes características. |
| Causalidad | relación entre causa y efecto. Generalmente identificados como variables. No hay que confundir causalidad con correlación. La correlación mide la similitud estructural  numérica entre dos variables. Normalmente la existencia de correlación es condición necesaria para la causalidad. |
| Coeficiente de correlación | estadístico que cuantifica la correlación. Sus valores están comprendidos entre –1 y 1 |
| Coeficiente de variación | medida de dispersión relativa. No tiene unidades y se calcula dividiendo la cuasi-desviación típica entre la media muestral. Se suele expresar en tanto por ciento. |
| Contraste bilateral | contraste de hipótesis en la que la hipótesis alternativa da opción a igualdad o superioridad. |
| Contraste de hipótesis | proceso estadístico que se sigue para la toma de decisiones a partir de la información de la muestra. Comparando el valor del estadístico experimental con el valor teórico rechazamos o no la hipótesis nula. |
| Correlación | concordancia entre dos variables según el sentido de la relación de estas en términos de aumento o disminución. |
| Datos censurados | en análisis de supervivencia son datos donde no se conoce el tiempo total hasta la aparición del fracaso / éxito bien porque el individuo se retiró del estudio, bien porque se acabó el estudio (datos con censura administrativa). Existen datos censurados por la izquierda y por la derecha. |
| Distribución de datos | en la realización de un experimento, corresponde a la recogida de los datos experimentales para cada individuo y cada variable. |
| Escala | la distribución de datos puede recogerse en distintas escalas: nominal, dicotómica,  discreta o continua. |
| Estimación | técnicas estadísticas que a partir de la información de la estadística descriptiva pretenden conocer cómo es la población en global. Existen técnicas de estimación  puntuales y por intervalos de confianza. |
| Factor de clasificación | variable que se usa para clasificar los datos experimentales en grupos. Los factores de clasificación son variables nominales. Cada factor de clasificación se compone de niveles. Así la variable “fumador” codificada como “nunca”, “exfumador”, “fumador actual” es un factor de clasificación con tres niveles. |
| Hipótesis | cualquier teoría que formule posibles líneas de trabajo experimental. Ver hipótesis nula y alternativa. |
| Imprecisión | error que se comete en la predicción |
| Modelo | intento matemático / estadístico para explicar una variable respuesta por medio de  una o más variables explicativas o factores. |
| Muestras | subgrupos de observaciones de la población de estudio. |
| Observación | sinónimo de caso, registro e indivíduo. |
| Parámetros | son valores desconocidos de características de una distribución teórica. El objetivo de la estadística es estimarlos bien dando un valor concreto, bien dado un intervalo confidencial. |
| Rango | diferencia entre el valor máximo y mínimo de una muestra o población. Solo es válido en variables continuas. |
| Sesgo | la diferencia entre el valor del parámetro y su valor esperado. También se utiliza en contraposición de aleatorio, así una muestra sesgada es no aleatoria. |
| Simetría | medida que refleja si los valores muéstrales se extienden o no de igual  forma a ambos lados de la media. |
| Transformaciones | cambios de escala con el propósito de conseguir linealidad, normalidad en los datos. |
| Valores numéricos | resultados de las variables para cada individuo en la muestra de estudio. Su naturaleza puede ser nominal, dicotómica, ordinal o continua. |
| Variable | objeto matemático que puede tomar diferentes valores. Generalmente asociado a propiedades o características de las unidades de la muestra. Lo contrario de variable es constante. |
| Variable aleatoria | variable cuyo resultado varía según la muestra según una distribución de probabilidad. |
| Variable continua | aquella que puede tomar una infinidad de valores, de forma que dados dos valores cualquiera, también pueda tomar cualquier valor entre dichos valores. |
| Variable discreta | variable que toma un número finito o infinito de valores, de forma que no cubre todos los posibles valores numéricos entre dos dados, en contraposición de las  continuas. |
| Variables | describen características en las observaciones realizadas. |

1. **REFERENCIAS BIBLIOGRÁFICAS:**

Devore, J. (2008). *Probabilidad y estadística para ingeniería y ciencias.* Cengage Learning. <https://intranetua.uantof.cl/facultades/csbasicas/matematicas/academicos/jreyes/DOCENCIA/APUNTES/APUNTES%20PDF/Probabilidad%20y%20Estadistica%20para%20Ingenieria%20y%20Ciencias%20-%20Jay%20Devore%20-%20Septima%20Edicion.pdf>

Orellana, L., D., y Sánchez, G., M. (2006). Técnicas de recolección de datos en entornos virtuales más usadas en la investigación cualitativa. *RIE - Revista de Investigación Educativa Asociación Interuniversitaria de Investigación Pedagógica Murcia,* *24*(1), 205-222. <https://www.redalyc.org/pdf/2833/283321886011.pdf>

Rustom, J., A. (2012). *Estadística descriptiva, probabilidad e inferencia. Una visión conceptual y aplicada.* Universidad de Chile. <https://repositorio.uchile.cl/bitstream/handle/2250/120284/Rustom_Antonio_Estadistica_descriptiva.pdf>

Tanvi, B., Gautam, S., & Vijay, M. (2020). *Determining Sufficient Volume of Data for Analysis with Statistical Framework*. Springer <https://www.springerprofessional.de/en/determining-sufficient-volume-of-data-for-analysis-with-statisti/18346230>

Torres, I., M., Paz, K., G., y Salazar, F. (s. f.). Métodos de recolección de datos para una investigación. *Boletín Electrónico No. 03*.<https://fgsalazar.net/LANDIVAR/ING-PRIMERO/boletin03/URL_03_BAS01.pdf>

1. **CONTROL DEL DOCUMENTO**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | Nombre | Cargo | Dependencia  *(Para el SENA indicar Regional y Centro de Formación)* | Fecha |
| Autor | Jaime Mauricio Peñaloza Trespalacios | Experto Técnico | Distrito Capital. Centro Electricidad Electrónica y Telecomunicaciones | Octubre de 2021. |
| [Leidy Carolina Arias Aguirre](mailto:leidyc.arias@misena.edu.co) | Diseñadora instruccional | Regional Distrito Capital- Centro de Diseño y Metrología | Octubre de 2021 |
| Carolina Coca Salazar | Revisora Metodológica y Pedagógica | Regional Distrito Capital- Centro de Diseño y Metrología | Diciembre de 2021 |
| José Gabriel Ortiz Abella | Corrector de estilo | Regional Distrito Capital - Centro de Diseño y Metrología | Abril del 2022. |
| Juan Gilberto Giraldo Cortés | Diseñador instruccional | Regional Tolima – Centro de Comercio y Servicios | Junio 2023 |
| María Inés Machado López | Metodóloga | Regional Tolima – Centro de Comercio y Servicios | Junio 2023 |

1. **CONTROL DE CAMBIOS**

**(Diligenciar únicamente si realiza ajustes a la Unidad Temática)**

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | Nombre | Cargo | Dependencia | Fecha | Razón del Cambio |
| Autor (es) |  |  |  |  |  |