



Módulo 2: Medidas de tendencia central y dispersión

≡ 1. Medidas descriptivas de centralización y dispersión

≡ 2. Distribución de datos y análisis aplicado de la variabilidad

≡ Referencias

1. Medidas descriptivas de centralización y dispersión

Las medidas descriptivas constituyen herramientas conceptuales que permiten resumir y caracterizar conjuntos de datos a partir de valores representativos. En el marco de la estadística descriptiva, estas medidas se orientan a sintetizar la información contenida en los datos, facilitando su análisis e interpretación sin recurrir a procedimientos inferenciales. Entre ellas se distinguen aquellas que describen la localización de los datos y aquellas que permiten analizar su dispersión o variabilidad (Zambrano Monserrate et al., 2023).

Las medidas de tendencia central buscan identificar un valor que represente el centro o punto de equilibrio de un conjunto de datos. La media aritmética se define como el cociente entre la suma de los valores observados y el número total de observaciones, y constituye una de las medidas más utilizadas debido a su simplicidad y propiedades algebraicas. Sin embargo, su valor puede verse afectado por observaciones extremas, lo

que limita su capacidad representativa en distribuciones asimétricas (INAP, 2020).

La mediana se define como el valor que divide al conjunto de datos ordenados en dos partes iguales, de modo que la mitad de las observaciones se sitúa por debajo y la otra mitad por encima de dicho valor. A diferencia de la media, la mediana no depende de la magnitud de los valores extremos, lo que la convierte en una medida adecuada cuando los datos presentan asimetrías o valores atípicos. Su interpretación se centra en la posición relativa de los datos dentro de la distribución (Zambrano Monserrate et al., 2023).

La moda corresponde al valor que aparece con mayor frecuencia en un conjunto de datos. Esta medida resulta particularmente relevante en el análisis de datos cualitativos, aunque también puede emplearse en datos cuantitativos. Su principal característica es que no siempre es única, ya que un conjunto puede presentar más de una moda o incluso carecer de ella. Por ello, su utilidad depende de la estructura de la distribución y del tipo de variable considerada (INAP, 2020).

Complementariamente, las medidas de dispersión permiten analizar el grado de variabilidad de los datos respecto de un valor central. La varianza se define como el promedio de los cuadrados de las desviaciones de los valores respecto de la media, y

proporciona una medida global de la dispersión del conjunto. Al elevar las desviaciones al cuadrado, esta medida enfatiza las diferencias más pronunciadas, lo que influye en su interpretación.

La desviación estándar surge como la raíz cuadrada de la varianza y se expresa en las mismas unidades que los datos originales, lo que facilita su interpretación. Esta medida indica, en promedio, cuánto se alejan los valores observados de la media, permitiendo evaluar la consistencia o heterogeneidad del conjunto de datos. Una desviación estándar pequeña sugiere valores concentrados, mientras que valores elevados indican mayor dispersión (INAP, 2020).

La consideración conjunta de las medidas de tendencia central y dispersión permite una descripción más completa de los datos. Mientras las primeras aportan información sobre la localización de los valores, las segundas permiten comprender la variabilidad interna del conjunto. Esta complementariedad resulta necesaria para evitar interpretaciones parciales y para construir una visión analítica más precisa del comportamiento de los datos en el marco de la estadística descriptiva (Zambrano Monserrate et al., 2023).

Medidas de tendencia central —

Las medidas de tendencia central se utilizan para describir la localización de los datos dentro de un conjunto, mediante la identificación de un valor representativo que sintetiza la información observada. Estas medidas permiten resumir grandes volúmenes de datos en un solo indicador, facilitando su análisis y comparación. En la estadística descriptiva, su empleo responde a la necesidad de caracterizar el comportamiento general de una variable cuantitativa a partir de su posición central (Zambrano Monserrate et al., 2023).

La media aritmética es una de las medidas de tendencia central más difundidas y se define como el resultado de dividir la suma de los valores observados por el número total de observaciones. Esta medida incorpora toda la información del conjunto de datos y posee propiedades matemáticas que la hacen especialmente útil para el análisis estadístico. Sin embargo, su valor depende directamente de la magnitud de cada observación, lo que puede afectar su representatividad en ciertos contextos (INAP, 2020).

Una característica relevante de la media es su sensibilidad frente a valores extremos. Cuando el conjunto de datos incluye observaciones muy alejadas del resto, la media puede desplazarse hacia dichos valores, ofreciendo una representación que no refleja adecuadamente la concentración principal de los datos. Esta propiedad debe ser considerada al interpretar resultados descriptivos, dado que puede generar conclusiones que no coincidan con la percepción general del conjunto (Zambrano Monserrate et al., 2023).

La mediana se define como el valor central de un conjunto de datos ordenados, de modo que separa las observaciones en dos grupos de igual tamaño. A diferencia de la media, la mediana se basa en la posición de los datos y no en su magnitud, lo que le otorga una mayor estabilidad frente a valores extremos. Esta característica la convierte en una medida adecuada cuando la distribución presenta asimetrías pronunciadas (INAP, 2020).

Desde una perspectiva conceptual, la mediana permite describir el punto a partir del cual la mitad de los datos se sitúa por debajo y la otra mitad por encima. Su interpretación no depende del valor específico de las observaciones extremas, sino del orden relativo de los datos dentro del conjunto. Por ello, resulta especialmente útil en contextos donde se busca una medida representativa de la posición central sin que intervengan distorsiones por valores atípicos (Zambrano Monserrate et al., 2023).

La moda se define como el valor que aparece con mayor frecuencia en un conjunto de datos. Esta medida puede aplicarse tanto a variables cualitativas como cuantitativas y se caracteriza por señalar el valor más recurrente dentro de la distribución. A diferencia de la media y la mediana, la moda no requiere operaciones aritméticas ni ordenamiento completo de los datos, sino un análisis de frecuencias (INAP, 2020).

Una particularidad de la moda es que no siempre es única. Un conjunto de datos puede presentar una sola moda, varias modas o carecer de ellas si todos los valores tienen la misma frecuencia. Esta característica limita su uso como único descriptor del centro, aunque puede aportar información relevante sobre la estructura de la distribución cuando se la considera junto con otras medidas .

La elección de una medida de tendencia central adecuada depende de la naturaleza de los datos y de la forma de su distribución. En conjuntos simétricos y sin valores extremos, la media, la mediana y la moda tienden a

coincidir o a presentar valores cercanos. En distribuciones asimétricas, en cambio, estas medidas pueden diferir notablemente, lo que exige una interpretación cuidadosa de cada indicador.

La consideración conjunta de la media, la mediana y la moda permite obtener una visión más completa de la localización de los datos. Cada medida aporta información distinta sobre el comportamiento del conjunto, y su análisis comparado favorece una descripción más precisa de la distribución. De este modo, las medidas de tendencia central cumplen una función descriptiva complementaria dentro del análisis estadístico (Zambrano Monserrate et al., 2023).

Medidas de dispersión y variabilidad —

Las medidas de dispersión permiten describir el grado de variabilidad existente dentro de un conjunto de datos, complementando la información aportada por las medidas de tendencia central. Mientras estas últimas identifican un valor representativo del centro de la distribución, las medidas de dispersión aportan información sobre la amplitud y heterogeneidad de los valores observados. Su análisis resulta necesario para comprender cómo se distribuyen los datos alrededor de un valor central.

La noción de variabilidad refiere a la existencia de diferencias entre los valores que componen un conjunto de datos. Cuando los valores se concentran en torno a un punto central, la variabilidad es reducida; en cambio, cuando los datos se encuentran más alejados entre sí, la variabilidad es mayor. Esta característica influye directamente en la interpretación de los resultados descriptivos y permite evaluar el grado de consistencia interna de la información observada.

Entre las medidas de dispersión más utilizadas se encuentra la varianza, definida como el promedio de los cuadrados de las desviaciones de cada valor respecto de la media. Al elevar las desviaciones al cuadrado, se eliminan los signos negativos y se otorga mayor peso a las diferencias más grandes. Esta medida proporciona una visión global de la dispersión del conjunto de datos, aunque su interpretación directa puede resultar menos intuitiva debido a que se expresa en unidades al cuadrado (Arreguín Samano, 2022).

La varianza permite cuantificar cuánto se apartan, en promedio, los valores observados del valor medio del conjunto. Un valor elevado de varianza indica que los datos presentan una dispersión considerable, mientras que valores bajos sugieren una mayor concentración en torno a la media. Esta medida resulta especialmente útil para comparar la variabilidad entre distintos conjuntos de datos que comparten la misma unidad de medida (INAP, 2020).

La desviación estándar se obtiene como la raíz cuadrada de la varianza y se expresa en las mismas unidades que los datos originales. Esta característica facilita su interpretación, ya que permite analizar la dispersión en términos comparables con los valores observados. Conceptualmente, la desviación estándar indica la distancia promedio entre los datos y la media del conjunto.

Desde una perspectiva analítica, la desviación estándar permite evaluar la consistencia de los datos. Valores bajos de desviación estándar indican que las observaciones se encuentran relativamente próximas entre sí y al valor medio, mientras que valores elevados reflejan una mayor heterogeneidad. Esta medida resulta particularmente relevante para interpretar la estabilidad de los datos en estudios descriptivos.

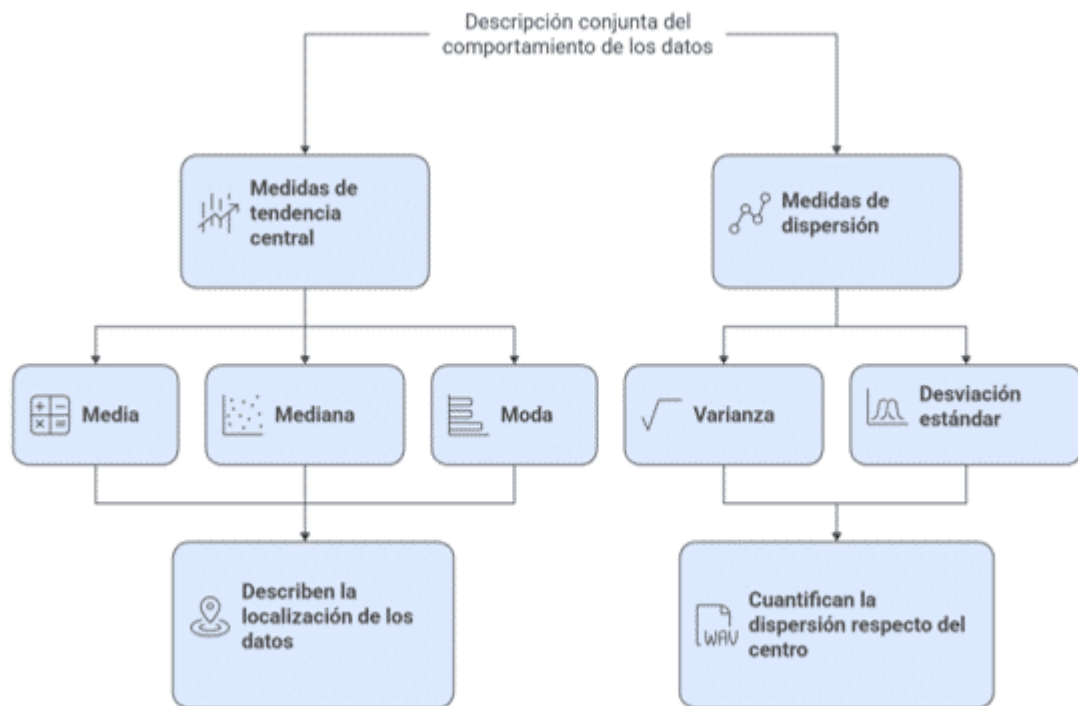
La relación entre varianza y desviación estándar es directa, dado que ambas miden la dispersión del conjunto a partir de las desviaciones

respecto de la media. No obstante, cada una cumple una función específica dentro del análisis descriptivo. Mientras la varianza se utiliza con frecuencia en desarrollos teóricos y comparativos, la desviación estándar es preferida para la interpretación y comunicación de resultados (Arreguín Samano, 2022).

La dispersión de los datos no puede analizarse de manera aislada, sino en relación con las medidas de tendencia central. Un conjunto de datos puede presentar un mismo valor medio y, sin embargo, diferir notablemente en su variabilidad. Por ello, considerar únicamente medidas de centralización puede conducir a interpretaciones incompletas del comportamiento de los datos (Zambrano Monserrate et al., 2023).

La incorporación de medidas de dispersión en el análisis descriptivo permite enriquecer la interpretación de la información y construir una representación más completa de los datos. Al analizar conjuntamente la localización y la variabilidad, se obtiene una visión más precisa del conjunto observado, lo que contribuye a una comprensión más rigurosa de la estructura y el comportamiento de los datos.

Figura 1. Medidas de tendencia central y dispersión en un conjunto de datos



Fuente: elaboración propia.

CONTINUAR

2. Distribución de datos y análisis aplicado de la variabilidad

La distribución de los datos constituye un elemento central en el análisis estadístico descriptivo, ya que permite observar cómo se organizan y agrupan los valores dentro de un conjunto. Analizar la distribución implica examinar la forma en que los datos se reparten a lo largo de los valores posibles, lo que aporta información relevante sobre su comportamiento general y su grado de concentración o dispersión (INAP, 2020).

La organización de los datos en distribuciones facilita su análisis al permitir una representación ordenada de la información. Mediante procedimientos de agrupamiento y conteo, los datos pueden resumirse de forma sistemática, lo que contribuye a identificar patrones, regularidades y diferencias internas. Esta organización resulta previa y necesaria para una interpretación adecuada de las medidas descriptivas calculadas posteriormente (Zambrano Monserrate et al., 2023).

La relación entre distribución de los datos y variabilidad es estrecha, dado que la forma que adopta la distribución influye directamente en la interpretación de las medidas de dispersión. Conjuntos de datos con distribuciones concentradas tienden a presentar menor variabilidad, mientras que distribuciones más extendidas o irregulares suelen asociarse con valores elevados de dispersión. Por ello, el análisis de la distribución permite contextualizar los indicadores de variabilidad (Arreguín Samano, 2022).

El análisis aplicado de la variabilidad requiere considerar simultáneamente la organización de los datos y las medidas descriptivas calculadas. La observación de cómo se distribuyen los valores en torno a un centro permite comprender mejor el significado de la varianza y la desviación estándar, evitando interpretaciones aisladas de estos indicadores. En este sentido, la variabilidad adquiere sentido analítico cuando se examina en relación con la estructura del conjunto de datos (INAP, 2020).

El uso de herramientas computacionales ha ampliado las posibilidades de análisis descriptivo al permitir procesar grandes volúmenes de datos de manera eficiente. Estas herramientas facilitan la organización de distribuciones y el cálculo de medidas de variabilidad, manteniendo la coherencia conceptual del análisis estadístico. Su incorporación no modifica los fundamentos teóricos del análisis, sino que optimiza los

procedimientos de cálculo y exploración de los datos (Python Software Foundation, 2024).

Desde una perspectiva analítica, la automatización del cálculo de distribuciones y medidas descriptivas exige una comprensión previa de los conceptos estadísticos involucrados. El uso de recursos computacionales requiere interpretar correctamente los resultados obtenidos y reconocer las relaciones entre distribución, centralización y dispersión. De este modo, la herramienta informática se integra como un medio para el análisis y no como un sustituto del razonamiento estadístico (Zambrano Monserrate et al., 2023).

La articulación entre distribución de datos y análisis de la variabilidad permite construir una descripción más completa y coherente de los conjuntos de datos observados. Al considerar conjuntamente la forma de la distribución y los indicadores de dispersión, se obtiene una visión analítica más precisa del comportamiento de los datos, fortaleciendo la interpretación dentro del marco de la estadística descriptiva (INAP, 2020).

Distribución de frecuencias y análisis de la variabilidad —

La distribución de frecuencias constituye un procedimiento conceptual mediante el cual los datos observados se organizan y agrupan para facilitar su análisis descriptivo. Este tipo de distribución permite ordenar los valores de una variable y cuantificar la cantidad de veces que cada valor o conjunto de valores aparece dentro del conjunto de datos. Su utilización favorece una visión estructurada de la información y permite identificar regularidades y diferencias internas (INAP, 2020).

Desde una perspectiva teórica, la distribución de frecuencias transforma un conjunto de datos desordenados en una representación sistemática que resume la información disponible. A través del conteo y la clasificación, los datos adquieren una forma organizada que facilita su interpretación. Este proceso resulta especialmente relevante cuando se trabaja con grandes volúmenes de datos, donde la observación directa de los valores individuales no permite reconocer patrones con claridad (Zambrano Monserrate et al., 2023).

Las distribuciones de frecuencias pueden construirse a partir de valores individuales o mediante la agrupación de datos en intervalos. En ambos casos, el objetivo es representar cómo se distribuyen los valores a lo largo del rango observado. La elección entre datos no agrupados o agrupados responde a criterios vinculados con la cantidad de observaciones y con la necesidad de síntesis de la información, sin alterar el significado estadístico de los datos.

El análisis de la variabilidad se vincula estrechamente con la distribución de frecuencias, dado que la forma que adopta la distribución aporta información sobre el grado de dispersión de los datos. Distribuciones concentradas en torno a ciertos valores indican menor variabilidad, mientras que distribuciones más extendidas o con valores dispersos

reflejan una mayor heterogeneidad. Esta relación permite interpretar las medidas de dispersión en un contexto más amplio.

La observación de la frecuencia con que aparecen los valores facilita la comprensión de cómo se alejan los datos respecto de un centro. Al analizar la distribución, es posible identificar si los valores se agrupan de manera uniforme o si presentan concentraciones y vacíos que influyen en la interpretación de la variabilidad. En este sentido, la distribución de frecuencias actúa como un complemento conceptual de las medidas de dispersión.

La variabilidad no se expresa únicamente a través de indicadores numéricos como la varianza o la desviación estándar, sino también mediante la forma general de la distribución de los datos. Una distribución simétrica tiende a presentar una dispersión equilibrada alrededor del centro, mientras que distribuciones asimétricas suelen asociarse con una dispersión desigual. Reconocer estas características contribuye a una interpretación más rigurosa del comportamiento de los datos (INAP, 2020).

El análisis conjunto de la distribución de frecuencias y la variabilidad permite evaluar la consistencia interna de los datos observados. Dos conjuntos de datos pueden presentar valores centrales similares, pero diferir notablemente en la forma de su distribución y en su grado de dispersión. Por ello, la sola consideración de medidas de centralización resulta insuficiente para describir adecuadamente el comportamiento de los datos (Zambrano Monserrate et al., 2023).

Desde el punto de vista descriptivo, la distribución de frecuencias facilita la comparación entre distintos conjuntos de datos al proporcionar una estructura común de análisis. Al observar cómo se distribuyen los valores y cómo varía su frecuencia, es posible identificar diferencias en la dispersión y en la concentración de los datos. Esta comparación se apoya en criterios conceptuales que no dependen del tamaño del conjunto analizado.

En conjunto, la organización de los datos mediante distribuciones de frecuencias y el análisis de la variabilidad permiten construir una descripción más completa y coherente de la información estadística. Esta articulación favorece una comprensión integral del comportamiento de los datos, al relacionar la forma de la distribución con el grado de dispersión, fortaleciendo así el análisis descriptivo (INAP, 2020).

Cálculo e interpretación de medidas descriptivas con herramientas computacionales

El uso de herramientas computacionales en el análisis estadístico descriptivo responde a la necesidad de procesar conjuntos de datos de mayor tamaño y complejidad de manera eficiente. Estas herramientas permiten automatizar el cálculo de medidas de tendencia central y dispersión, reduciendo errores operativos y facilitando la exploración sistemática de la información. Su incorporación se inscribe dentro del marco teórico de la estadística descriptiva, sin modificar los principios conceptuales que sustentan el análisis (Zambrano Monserrate et al., 2023).

El cálculo computacional de medidas descriptivas se apoya en algoritmos que reproducen las definiciones matemáticas de indicadores como la media, la mediana, la moda, la varianza y la desviación estándar. En este sentido, las herramientas informáticas no sustituyen el razonamiento estadístico, sino que ejecutan de forma automatizada operaciones previamente definidas desde el punto de vista teórico. La comprensión de estas definiciones resulta necesaria para interpretar adecuadamente los resultados obtenidos (INAP, 2020).

Las hojas de cálculo, como Excel, constituyen uno de los recursos más difundidos para el análisis descriptivo de datos. Estas herramientas permiten organizar la información en estructuras tabulares y aplicar

funciones predefinidas para el cálculo de medidas estadísticas. Desde una perspectiva conceptual, su utilidad radica en la posibilidad de vincular la organización de los datos con el análisis de la variabilidad y la centralización, favoreciendo una lectura integrada de los resultados.

Por su parte, los lenguajes de programación orientados al análisis de datos, como Python, incorporan bibliotecas especializadas que amplían las posibilidades del análisis estadístico descriptivo. Librerías como pandas y NumPy permiten trabajar con estructuras de datos complejas y aplicar funciones estadísticas de manera flexible. Estas herramientas facilitan el manejo de grandes volúmenes de datos y la repetición sistemática de análisis bajo criterios consistentes.

Desde el punto de vista analítico, el uso de herramientas computacionales exige interpretar los resultados en función del contexto estadístico en el que se inscriben. El valor de una media o de una desviación estándar calculada mediante software adquiere significado únicamente cuando se analiza en relación con la distribución de los datos y con las características de la variable observada. Por ello, el cálculo automatizado debe ir acompañado de una interpretación conceptualmente fundamentada (Zambrano Monserrate et al., 2023).

La comparación de resultados obtenidos mediante distintas herramientas computacionales permite reconocer la consistencia de las medidas descriptivas. Dado que las funciones implementadas se basan en las mismas definiciones matemáticas, los valores calculados deben coincidir, salvo diferencias derivadas de criterios específicos de cálculo. Este aspecto refuerza la importancia de comprender los supuestos teóricos subyacentes al uso de cada herramienta (INAP, 2020).

El análisis computacional de la variabilidad facilita la exploración de cómo se dispersan los datos alrededor de un valor central. La obtención rápida de indicadores como la varianza y la desviación estándar permite evaluar

distintos conjuntos de datos y comparar su grado de heterogeneidad. Sin embargo, estos indicadores conservan su carácter descriptivo y deben interpretarse en relación con la estructura de la distribución observada.

La integración de herramientas computacionales en el análisis estadístico descriptivo favorece una aproximación sistemática y reproducible al estudio de los datos. La posibilidad de repetir cálculos y aplicar los mismos criterios analíticos a distintos conjuntos de información fortalece la coherencia del análisis y contribuye a una comprensión más profunda de las medidas descriptivas y su significado estadístico (Zambrano Monserrate et al., 2023).

En conjunto, el cálculo e interpretación de medidas descriptivas mediante herramientas computacionales se inscribe dentro del marco teórico de la estadística descriptiva como un recurso de apoyo al análisis. Su valor reside en la capacidad de articular automatización y razonamiento estadístico, permitiendo describir con mayor precisión la centralización y la variabilidad de los datos sin perder de vista los fundamentos conceptuales del análisis .

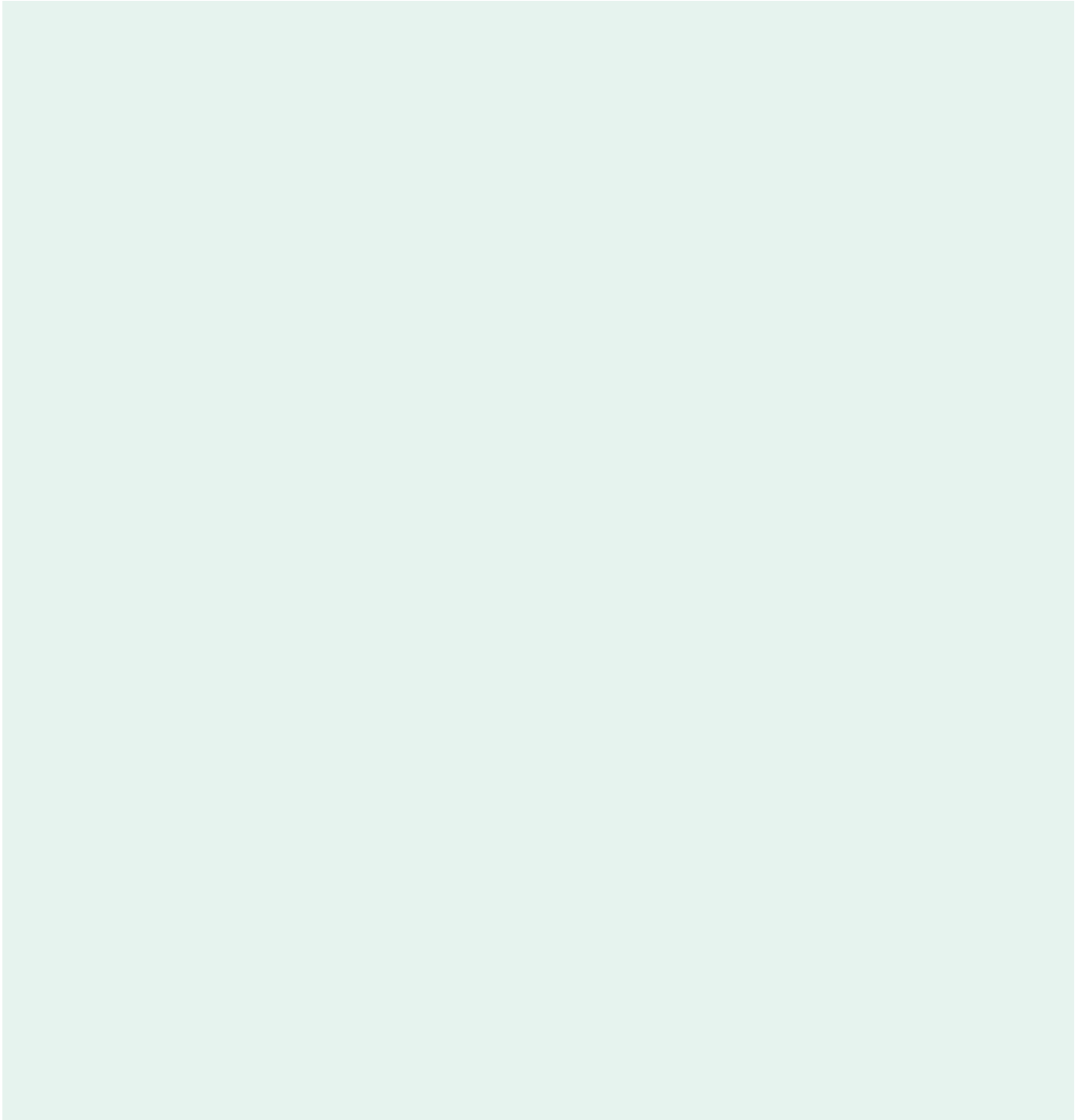
Figura 2. Relación entre distribución de datos, variabilidad y herramientas computacionales.



Fuente: elaboración propia.

CONTINUAR

Referencias



Arreguín Samano, M. (2022). *Estadística para ingeniería*. Editorial Universidad de Bogotá. <https://biblioteca.ciencialatina.org/wp-content/uploads/2023/05/Estadistica-para-Ingenieria.pdf>

Instituto Nacional de la Administración Pública. (2020). *Manual de estadística básica para no especialistas*. Secretaría de Gestión y Empleo Público.

https://www.argentina.gob.ar/sites/default/files/cuinap_13_2020_0.pdf

Python Software Foundation. (2024). *statistics — Funciones de estadística matemática*.

<https://docs.python.org/es/3/library/statistics.html>

Zambrano Monserrate, M., Berrús Zhumi, A., & Goncalves Guillén, G. (2023). *Principios de estadística*. Universidad Espíritu Santo.

<https://uees.edu.ec/descargas/libros/2023/principios-de-estadistica.pdf>

CONTINUAR