

Capítulo 3

<https://doi.org/10.62486/978-628-97230-1-4.ch03>

Ejercicios prácticos: trabajando con datos no agrupados

Ahora que ya conoces los conceptos fundamentales de la estadística descriptiva, es momento de poner manos a la obra. En este capítulo trabajaremos con datos no agrupados, es decir, con listas de valores individuales que no han sido organizados en intervalos ni clases. Estos datos nos permiten aplicar de manera directa las medidas de tendencia central y dispersión.

Antes de comenzar, es importante recordar que los datos no agrupados reflejan observaciones simples, tal como fueron recolectadas. En muchas situaciones reales, especialmente cuando se trata de muestras pequeñas o de análisis rápidos, este tipo de datos es el más común. Su análisis puede hacerse sin recurrir a agrupamientos, y aunque puede parecer más laborioso cuando el conjunto es grande, también es más preciso en ciertos aspectos.

Imaginemos el primer caso. En una pequeña fábrica de artesanías, se utiliza una aleación compuesta por dos materiales: A y B. Después de una jornada de producción, se realiza un muestreo sobre la proporción del componente A presente en cada pieza. El resultado fue una lista de 26 mediciones en orden creciente: 2.1, 2.3, 2.5, 2.6, 2.6, 2.7, 2.7, 2.8, 3.0, 3.1, 3.2, 3.3, 3.3, 3.4, 3.4, 3.6, 3.6, 3.6, 3.6, 3.7, 4.4, 4.6, 4.7, 4.8, 5.4 y 10.0.

Lo primero que podríamos preguntarnos es: ¿cuál es el valor promedio de estas mediciones? Para responder, calculamos la media aritmética, que consiste en sumar todos los valores y dividir el resultado entre la cantidad total de observaciones. Al realizar esta operación, obtenemos una media aproximada de 3.6. Esto nos indica que, en promedio, la proporción del componente A en las piezas analizadas gira en torno a ese valor.

Sin embargo, si analizamos más a fondo, notamos algo interesante: el último valor es 10.0, muy superior al resto. Esto es lo que se conoce como un **valor atípico**, y puede influir en el resultado de la media. Para contrarrestar este efecto, podemos calcular la mediana. Dado que hay 26 datos, la mediana será el promedio de los valores 13 y 14, que corresponden a 3.3 y 3.4. La mediana, por tanto, es 3.35. Este resultado nos ofrece una visión alternativa del centro de la distribución, menos afectada por los extremos.

La moda, en este conjunto, también nos da información útil. Observamos que el valor 3.6 aparece cuatro veces, más que cualquier otro. Eso convierte a 3.6 en la moda, el valor más frecuente del conjunto. En este caso, tanto la media como la moda coinciden, lo que sugiere una cierta estabilidad en torno a ese número, aunque el valor extremo de 10.0 sigue destacando.

Para entender mejor la variabilidad de los datos, calculamos la desviación estándar, que nos indica cuánto se alejan, en promedio, los valores respecto a la media. En este conjunto, la desviación estándar es relativamente baja, excepto por la presencia del valor 10.0, que aumenta la dispersión general. El rango, que es la diferencia entre el mayor y el menor valor, también refleja esta variabilidad: 10.0 menos 2.1 nos da un rango de 7.9. Esto nos dice que hay una amplitud considerable entre los extremos de la muestra.

Pasemos ahora a un segundo ejemplo. En una industria ferroviaria, se analizaron los tiempos de respuesta de 26 trabajadores durante un ejercicio práctico. Los valores, sin ningún orden específico, fueron: 388, 357, 358, 364, 376, 423, 324, 395, 403, 372, 372, 371, 365, 365, 364, 324, 340, 393, 392, 370, 373, 358, 357, 404, 333 y 397.

Al ordenar los datos de menor a mayor, podemos calcular la media. Esta resulta ser aproximadamente 371.5. La mediana, que al ser una muestra par se obtiene del promedio entre los valores 13 y 14, es 365. Al comparar la media y la mediana,

observamos que están relativamente cerca, lo que sugiere que la distribución no está demasiado sesgada. La moda, en este caso, aparece en los valores 357, 358 y 365, todos repetidos dos veces, lo que indica una distribución multimodal, aunque poco marcada.

El rango de este conjunto es de 423 menos 324, es decir, 99 unidades. A simple vista, esta diferencia puede parecer amplia, pero al revisar la desviación estándar —alrededor de 24.8— observamos que la mayoría de los datos están razonablemente agrupados en torno a la media. Esta dispersión moderada puede deberse a la naturaleza de la actividad evaluada, donde las condiciones laborales y físicas influyen de manera parecida en todos los trabajadores.

En ambos ejercicios, además de los cálculos, es fundamental realizar un análisis de interpretación. La estadística no termina en los números. Por el contrario, comienza a ser útil cuando esos números nos dicen algo sobre la realidad. En el primer caso, podríamos concluir que la producción de la aleación es bastante estable, con una excepción notable que tal vez requiere revisión. En el segundo, el rendimiento de los trabajadores es bastante homogéneo, lo cual puede interpretarse como una buena señal de consistencia operativa.

Este proceso de observación, cálculo e interpretación es el núcleo de la estadística descriptiva. A medida que avances, notarás que cada paso aporta una capa más de comprensión. Y lo más importante: desarrollarás una habilidad crítica para analizar datos que no solo te servirá en la escuela, sino en cualquier entorno donde la información sea clave para tomar decisiones.

Ejercicio 1: análisis completo de datos no agrupados

Uno de los objetivos principales de la estadística descriptiva es poder analizar conjuntos de datos reales, interpretarlos y extraer conclusiones útiles. A continuación, trabajaremos con un caso tomado directamente de un contexto

productivo, que nos permitirá aplicar las medidas de tendencia central, de dispersión y de posición, así como interpretar su significado.

Imaginemos que estamos en una línea de producción de una artesanía mexicana donde se utiliza una aleación compuesta por dos componentes diferentes: A y B. De acuerdo con un muestreo realizado, se determinó que el componente A presenta mayor porcentaje en las muestras analizadas. La tabla que sigue muestra los datos obtenidos, ya ordenados de menor a mayor:

2.1 2.3 2.5 2.6 2.6 2.7 2.7 2.8 3.0 3.1

3.2 3.3 3.3 3.4 3.4 3.6 3.6 3.6 3.6 3.7

4.4 4.6 4.7 4.8 5.4 10.0

Contamos con un total de 26 datos.

Media (promedio): La media se obtiene sumando todos los valores y dividiendo entre el número total de observaciones.

Suma total = 92.5

Media $\approx 92.5 / 26 = 3.56$

Este valor representa el punto central teórico del conjunto. Sin embargo, conviene observar que el último dato (10.0) es notablemente mayor que los demás. Este valor extremo influye en la media, elevándola ligeramente. Por eso, conviene considerar también otras medidas.

Mediana: Al haber 26 datos, la mediana se calcula como el promedio entre el dato 13 y el 14.

Datos 13 y 14 = 3.3 y 3.4

$$\text{Mediana} = (3.3 + 3.4) / 2 = 3.35$$

La mediana nos indica que la mitad de los datos están por debajo de 3.35 y la otra mitad por encima. Al no estar afectada por los valores extremos, suele ser una medida más robusta de la tendencia central cuando hay datos atípicos.

Moda: El valor que más se repite es 3.6, el cual aparece cuatro veces, más que cualquier otro.

$$\text{Moda} = 3.6$$

Esto sugiere que, aunque la media y la mediana están ligeramente por debajo, el valor más representativo en términos de frecuencia es 3.6.

Rango: $\text{Rango} = \text{valor máximo} - \text{valor mínimo} = 10.0 - 2.1 = 7.9$

Esto nos indica la amplitud completa de los datos.

Desviación estándar: Después de aplicar la fórmula correspondiente, encontramos una desviación estándar de aproximadamente 1.60, lo que significa que, en promedio, los datos se separan de la media por 1.6 unidades. Este valor es moderado, pero está claramente influido por el dato extremo (10.0), lo que sugiere una cierta dispersión irregular.

Cuartiles:

- Q1 (primer cuartil) ≈ 2.85
- Q2 (mediana) $= 3.35$
- Q3 (tercer cuartil) ≈ 3.95

El rango intercuartílico ($RIQ = Q3 - Q1$) es ≈ 1.10 . Esto nos indica que el 50% central de los datos se encuentra contenido dentro de un intervalo relativamente estrecho.

Diagrama de caja y bigote: Este diagrama permite visualizar rápidamente la dispersión de los datos, los cuartiles y posibles valores atípicos. En este caso, el valor 10.0 aparece como un outlier evidente, separado del resto del grupo, cuya concentración principal está entre los valores de 2.8 y 4.0.

Gráfica lineal: Al representar cada valor de la serie sobre una línea, se observa una progresión bastante suave, con una gran concentración de valores entre 2.5 y 4.0. El salto repentino al valor 10.0 rompe el patrón, lo que sugiere que puede ser un error de medición o una observación excepcional.

Análisis general: Los datos analizados reflejan una distribución bastante simétrica y estable entre los valores 2.5 y 4.0, con una media (3.56) y moda (3.6) casi idénticas. La presencia de un valor extremo (10.0) eleva la media y la desviación estándar, pero no afecta la mediana ni los cuartiles de forma significativa. Esto nos lleva a concluir que, en general, la proporción del componente A es bastante constante en la mayoría de las piezas, aunque debe analizarse el origen del dato más alto, ya que podría representar un lote defectuoso o un proceso irregular.

Este ejercicio demuestra cómo una serie de datos puede ser analizada con herramientas sencillas para obtener una visión clara del comportamiento de un fenómeno. Cada medida aporta una perspectiva distinta, y juntas permiten tomar decisiones informadas basadas en evidencia.

En el siguiente capítulo exploraremos cómo cambia el análisis cuando los datos se agrupan. Esto ocurre cuando el volumen de información crece tanto que conviene organizar los datos en intervalos para simplificar su manejo. Aunque el principio es el mismo —describir, entender, interpretar—, las herramientas y los cálculos se ajustan para adaptarse al nuevo formato.

Ejercicio 2: Vida útil de brocas fresadoras

Este caso proviene de un escenario ficticio, en el que se analizaron los tiempos de vida de microbrocas utilizadas para fresar agujeros en una aleación de latón. Se obtuvieron los siguientes datos (ordenados de menor a mayor):

11, 14, 20, 23, 59, 61, 65, 67, 81, 84, 85, 89, 105, 105, 112, 118, 161, 168, 184, 206

Son 20 valores en total, lo que permite un análisis estadístico sencillo y completo.

Media (promedio): Al sumar todos los valores, obtenemos un total de 2,256.

$$\text{Media} \approx 2256 / 20 = 112.8$$

Esto indica que, en promedio, las brocas duran aproximadamente 113 unidades de tiempo (según la medida empleada en el estudio).

Mediana: Con 20 datos, la mediana se obtiene promediando los valores 10 y 11.

$$\text{Valores 10 y 11} = 84 \text{ y } 85$$

$$\text{Mediana} = (84 + 85) / 2 = 84.5$$

Aquí observamos una diferencia importante: la mediana es bastante menor que la media. Eso ocurre porque los últimos valores (161, 168, 184, 206) son muy altos

y elevan la media. Esta situación sugiere una distribución asimétrica hacia la derecha, donde unos pocos valores altos alteran el centro aritmético del conjunto.

Moda: Hay un valor que se repite: 105.
Moda = 105

Aunque solo aparece dos veces, es el único valor duplicado.

Rango: $206 - 11 = 195$

El rango es muy amplio, lo que refleja una gran variabilidad en la duración de las brocas. Esto puede deberse a diferencias en materiales, condiciones de uso, o incluso defectos de fabricación.

Desviación estándar: El cálculo nos da una desviación estándar aproximada de 49.6, lo que indica una alta dispersión respecto a la media. Los valores más altos están muy separados del centro.

Cuartiles:

- $Q1 \approx 65$
- $Q2 = 84.5$ (mediana)
- $Q3 \approx 118$

$$RIQ = Q3 - Q1 = 53$$

El rango intercuartílico nos muestra que la mitad central de los datos está bastante agrupada, a pesar de los valores extremos.

Análisis: La distribución es asimétrica positiva (hacia la derecha). La mayoría de las brocas dura menos de 120 unidades de tiempo, pero hay casos que alcanzan duraciones notablemente superiores. Esto puede deberse a diferencias en calidad o condiciones de uso. Para fines de análisis de calidad, sería útil revisar el motivo de estos valores altos: ¿son mejoras o excepciones?

Ejercicio 3: Comparación entre dos compañías

En este caso, se evaluaron dos conjuntos de varillas metálicas provistas por dos empresas diferentes. Se fabricaron diez resortes con cada lote y se midieron sus niveles de flexibilidad. Los datos registrados fueron:

Compañía A: 9.3, 8.8, 6.8, 8.7, 8.5, 6.7, 8.0, 6.5, 9.2, 7.0

Compañía B: 11.0, 9.8, 9.9, 10.2, 10.1, 9.7, 11.0, 11.1, 10.2, 9.6

Nuestro objetivo es comparar ambas series con base en medidas de tendencia central y dispersión.

Compañía A:

Suma = 79.5

Media ≈ 7.95

Mediana = 8.15

Moda = no hay valores repetidos

Rango = $9.3 - 6.5 = 2.8$

Desviación estándar ≈ 0.96

Compañía B:

Suma = 103.6

Media ≈ 10.36

Mediana = 10.1

Moda = 11.0 (repetido)

Rango = $11.1 - 9.6 = 1.5$

Desviación estándar ≈ 0.51

Análisis comparativo: La compañía B presenta una media significativamente más alta y una desviación estándar más baja. Esto significa que no solo sus productos son más flexibles en promedio, sino que también lo son de forma más consistente. Mientras que los datos de la compañía A están más dispersos, los de la compañía B están agrupados más estrechamente en torno al promedio.

En este caso, un análisis estadístico bien realizado permite tomar una decisión fundamentada: si se busca mayor flexibilidad y estabilidad en la producción, los datos apuntan claramente hacia la Compañía B como la opción más confiable.

A través de estos tres ejercicios, hemos visto cómo aplicar los conceptos básicos de estadística descriptiva a datos no agrupados. Hemos aprendido a calcular la media, mediana, moda, rango, desviación estándar y cuartiles, y lo más importante: hemos interpretado lo que esos valores nos dicen en contextos reales.

En el siguiente capítulo abordaremos el análisis de datos agrupados. Veremos que, aunque la lógica es similar, las herramientas y fórmulas se ajustan para trabajar con grupos de datos resumidos en intervalos. Es un paso más en tu camino hacia la comprensión total de la estadística.

Derechos de Autor (Copyright) Guillermo Alejandro Zaragoza Alvarado 2025 ©

Este texto está protegido por una licencia Creative Commons 4.0.



Usted es libre de compartir, copiar y redistribuir el material en cualquier medio o formato, así como de adaptarlo, remezclarlo, transformarlo y crear a partir de él para cualquier propósito, incluso con fines comerciales. Sin embargo, debe cumplir con la condición de atribución, lo que significa que debe otorgar el crédito correspondiente a la obra original de manera adecuada, proporcionar un enlace a la licencia e indicar si se han realizado modificaciones. Puede hacerlo en cualquier formato razonable, pero no de manera que sugiera que cuenta con el respaldo del licenciante o que recibe algún beneficio por el uso de la obra.

[Resumen de licencia](#) – [Texto completo de la licencia](#)

ISBN: 978-628-97230-1-4

DOI: 10.62486/978-628-97230-1-4.ch03