Facultad de Ingeniería

Asignatura: Probabilidad y Estadística

Estadística Descriptiva

Docente: Ing. Alberto Acosta López

Grupo: 020-84

Estudiantes:

Juan Felipe Wilches Gómez	Código: 20231020137
Tomás Arévalo Montes	Código: 20232020267
Daniel Vargas Arias	Código: 20232020103

Bogotá D.C.

2025

Resumen

El presente informe aborda los fundamentos de la estadística descriptiva, disciplina esencial para la recopilación, organización, análisis e interpretación de datos con el fin de facilitar la comprensión de fenómenos y apoyar la toma de decisiones informadas (Pareja et al., 2025). Se estudian los principales tipos de variables y su clasificación, las medidas de tendencia central, de dispersión y de posición, así como las representaciones gráficas más utilizadas para describir conjuntos de datos. A través de ejemplos prácticos, un enfoque recomendado por (Triola, 2018) se evidencia cómo la estadística descriptiva permite sintetizar grandes volúmenes de información, identificar patrones y establecer bases sólidas para el análisis inferencial. En conjunto, este trabajo resalta la importancia de la estadística descriptiva como herramienta fundamental en los procesos científicos, tecnológicos y de gestión en ingeniería.

1. Introducción

La estadística descriptiva es la disciplina que recolecta, organiza, resume y presenta datos con el propósito de facilitar su interpretación y la toma de decisiones (Pareja, 2025). Su origen se remonta a las prácticas censales de civilizaciones antiguas y al desarrollo de la *statistik* en el siglo XVIII (López, 2019). En el siglo XIX y XX, científicos como Quetelet, Pearson y Galton desarrollaron conceptos de tendencia central, correlación, regresión y diseño experimental que formalizaron la disciplina (El nacimiento de la estadística moderna. Francis Galton y Karl Pearson (2020) (Akal, 2020).

En el estado del arte actual, la estadística descriptiva se integra con la ciencia de datos y la inteligencia artificial: se automatiza el EDA (*exploratory data analysis*)(IBM, 2025), se aplican técnicas robustas para *big data* e IoT, y se enfatiza la ética y la interpretabilidad en la automatización de decisiones.

2. Importancia y aplicaciones en ingeniería de sistemas

La estadística descriptiva es una herramienta fundamental en todas las ramas de la ingeniería, ya que proporciona métodos para organizar, resumir y presentar datos de procesos experimentos o sistemas.

En la ingeniería de sistemas es especialmente útil a la hora de realizar el front end de una aplicación de software ya que los datos numéricos y algebraicos no se pueden mostrar crudos a los usuarios o clientes, hablando en el caso de un software empresarial. Estos necesitan ser interpretados primero mediante la estadística descriptiva para poder darlos a entender al público.

Un ingeniero de sistemas debe garantizar que las aplicaciones y la infraestructura funcionen de manera eficiente. La estadística descriptiva es la base para lograrlo.

- **Caracterización de la carga:** Al utilizar métricas descriptivas (media, mediana, percentiles, desviación estándar), los ingenieros pueden analizar el tráfico de red, el tiempo de respuesta de los servidores o el uso de recursos (CPU, memoria). Por ejemplo, calcular el percentil

95 del tiempo de respuesta no solo ofrece un promedio, sino que describe el peor caso que experimenta la mayoría de los usuarios, lo cual es crucial para establecer acuerdos de nivel de servicio (SLA).

- **Identificación de cuellos de botella:** La visualización de datos mediante histogramas o diagramas de caja permite identificar rápidamente dónde se concentra la mayor latencia o variabilidad en los tiempos de procesamiento. Esta información es vital para enfocar los esfuerzos de optimización donde tendrán el mayor impacto en el rendimiento global del sistema.

3. Avances recientes y estado del arte

Los avances en la estadística descriptiva destacan por ser aplicaciones de esta en la tecnología contemporánea, siendo estos avances en áreas como la automatización, el Big Data y la integración con la inteligencia artificial.

La estadística descriptiva automatizada impulsada por los DS-Agents (Agentes de ciencias de datos) o la inteligencia artificial, véase el caso de DS-STAR, un agente de ciencia de datos desarrollado por Google Research que puede analizar automáticamente conjuntos de datos complejos, identificar variables, detectar distribuciones y generar un resumen de texto y gráfico, reduciendo el tiempo invertido en fases del *pipeline* de un proyecto como la fase de exploración de datos. Según Google Research (2025), "DS-STAR es un agente de ciencia de datos de última generación cuya versatilidad se demuestra por su capacidad para automatizar una variedad de tareas — desde análisis estadísticos hasta visualización y manipulación de datos — a través de diversos tipos de datos, culminando en un rendimiento destacado en el famoso benchmark DABStep." (Yoon & Nam, 2025).

Algoritmos avanzados son capaces de calcular límites estadísticos de un proceso y señalar automáticamente los valores atípicos o cambios bruscos en la distribución; también detectan anormalidades y ahorran tiempo y recursos en procesos tediosos.

Otros avances en la estadística descriptiva se presentan como herramientas para la narrativa de datos (*data storytelling*, del inglés), ayudando a crear coherentemente guías de usuario para comunicar los datos. Flourish, una plataforma desarrollada por Canva, se destaca por permitir a los usuarios "crear narrativas visuales interactivas y atractivas sin necesidad de conocimientos de programación" (Flourish, s. f.). Este software está especializado en generar gráficos, mapas de calor y visualizaciones dinámicas que facilitan la comprensión de conjuntos de datos complejos.

Asimismo, herramientas como Tableau, que posibilita la creación de *dashboards* interactivos enlazados con múltiples gráficas, representan ejemplos destacados de esta tendencia hacia la automatización y la accesibilidad en la comunicación de datos. Según Tom Perry, director sénior de datos, información e integración de Elsevier, "Tableau es un programa para cambiar la empresa, la cultura y la forma de analizar los datos y tomar decisiones. A medida que avanzamos en nuestro camino y le mostramos al personal lo fácil que era obtener información por su cuenta, tuvimos epifanías muy importantes. Ahora podemos dar un paso atrás y brindarle apoyo en su aprendizaje en lugar de tener que centralizar todo" lo que refleja la evolución de las plataformas analíticas hacia sistemas más inteligentes y autónomos, capaces de asistir al usuario en todo el proceso de análisis de datos, desde la exploración hasta la toma de decisiones (Tableu, 2021).

4. Desarrollo

4.1 Conceptos básicos

La estadística descriptiva inicia con la correcta identificación y clasificación de la información. Para ello, se distinguen conceptos fundamentales que permiten comprender cómo se estructuran y analizan los datos antes de aplicar técnicas estadísticas.

Población, muestra, variable y dato

Población: conjunto total de elementos u objetos de estudio. **Muestra:** subconjunto representativo de la población, del cual se obtienen los datos. **Variable:** característica que puede tomar diferentes valores entre los elementos de una población. **Dato:** valor específico que toma una variable en un elemento de la muestra.

Estos conceptos permiten definir el alcance del análisis y determinar las técnicas adecuadas para la representación y tratamiento de los datos.

Tipos de variables

- **Cualitativas:**
  - Nominales: categorías sin orden (ej: tipo de sangre, navegador usado).
  - Ordinales: categorías con jerarquía (ej: nivel de satisfacción).
  - Dicotómicas: solo dos categorías (ej: sexo, moneda).
- **Cuantitativas:**
  - Discretas: valores enteros (ej: número de solicitudes por segundo).
  - Continuas: valores reales (ej: tiempo de respuesta en milisegundos).

Escalas de medición

Las variables pueden medirse en escalas nominal, ordinal, de intervalo y de razón. Estas determinan qué operaciones estadísticas pueden aplicarse y qué visualizaciones son apropiadas. Tenemos el ejemplo de una competencia con los tres primeros puestos:

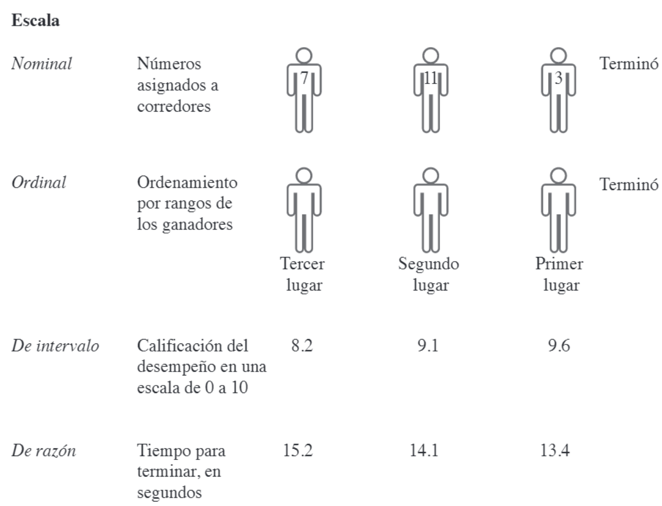


Figure 1: Gráfica descriptiva de las escalas

4.2 Presentación y transformación de datos mediante ejemplos

Una parte fundamental de la estadística descriptiva consiste en organizar los datos para facilitar su interpretación. A continuación se integran los conceptos anteriores con ejemplos prácticos y visualizaciones que ilustran su utilidad.

Datos sin agrupar

Los datos sin agrupar corresponden a valores individuales tal como se recogen en la muestra. Se emplean para cálculos exactos de media, mediana, moda y percentiles.

Graficos para trabajar Datos sin agrupar

Cuando trabajamos con datos sin agrupar, las mejores gráficas para analizarlos son:

- **Histograma con líneas de media, mediana, moda y percentiles.**

  El histograma es una de las representaciones más completas para el análisis descriptivo de datos sin agrupar, ya que permite:

  – Visualizar la forma de la distribución (simetría, sesgo, curtosis).

  - Detectar sesgos o acumulaciones de valores.
  - Identificar valores frecuentes con claridad.
  - Destacar medidas estadísticas (media, mediana, moda y percentiles) mediante líneas verticales.

- **Boxplot (diagrama de caja).**

  El boxplot es ideal para comprender de forma rápida la dispersión de los datos. Permite observar:

  - Los percentiles principales (Q1, Q2, Q3).
  - La mediana de los datos.
  - El rango intercuartílico (IQR).
  - La presencia de valores atípicos.

  Aunque no muestra la forma detallada de la distribución, es altamente efectivo para resumir la variabilidad y detectar anomalías en los datos.

- **Diagramas de dispersión.**

  Son útiles para visualizar la posición individual de cada dato dentro del conjunto. Sus ventajas principales son:

  - Mostrar cada observación de manera aislada.
  - Facilitar la identificación de patrones o concentraciones.
  - Resaltar valores atípicos o atípicamente altos/bajos.

  Aunque no representan la frecuencia, ayudan a entender la distribución espacial de los datos.

- **Curva KDE (Estimación de Densidad Kernel) con todas las medidas marcadas.**

  La curva KDE permite visualizar la distribución suavizada de los datos sin agrupar. Resulta especialmente útil para:

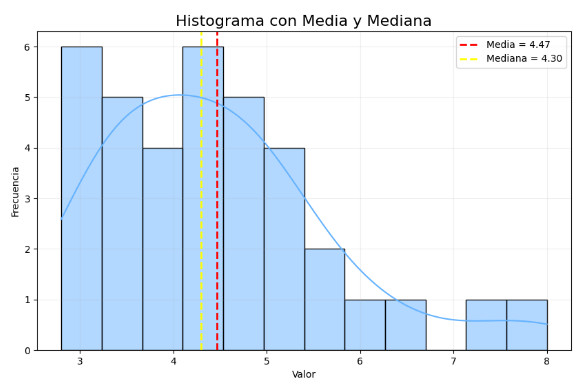
– Identificar la forma de la distribución de manera continua.

– Comparar visualmente zonas de mayor y menor densidad.

– Superponer líneas de media, mediana, moda y percentiles para un análisis completo.

Combinada con líneas verticales de medidas estadísticas, se convierte en una de las gráficas más informativas visualmente.

Ejemplo aplicado: Un equipo de backend registra 50 tiempos de respuesta para evaluar degradación bajo pico. Queremos medidas exactas (media, mediana, percentiles, moda), histograma para ver estructura. Datos: 2.8, 3.0, 3.1, 3.2, 3.2, 3.2, 3.3, 3.4, 3.5, 3.6, 3.6, 3.8, 3.8, 4.0, 4.0, 4.1, 4.1, 4.2, 4.4, 4.5, 4.5, 4.6, 4.7, 4.8, 4.8, 4.9, 5.0, 5.0, 5.2, 5.3, 5.5, 5.7, 6.0, 6.5, 7.5, 8.0



Interpretación del histograma con media y mediana

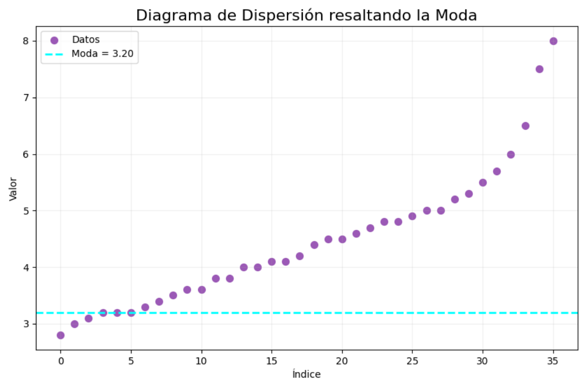
El histograma revela información clave sobre los datos:

- **Forma de la distribución:** permite observar si los tiempos están concentrados, dispersos, distribuidos simétricamente o si presentan asimetría hacia la derecha, algo común en tiempos de respuesta bajo congestión.

- **Línea de la media:** Representa el promedio general y es sensible a valores extremos. Si la media está a la derecha de la mediana, sugiere una distribución asimétrica positiva.

- **Línea de la mediana:** Señala el valor central y no se ve afectada por valores extremos. Si es mucho menor que la media, implica presencia de outliers altos.

Conclusión: El histograma permite identificar si la degradación del sistema es general o causada por picos aislados.



Interpretación del diagrama de dispersión con moda

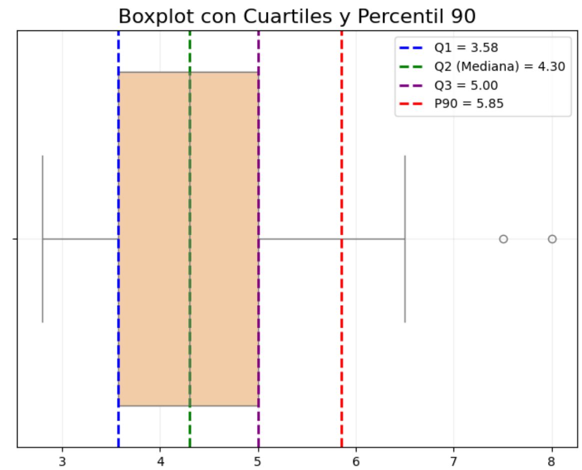
En esta gráfica:

- **Puntos individuales:** muestran cada tiempo registrado, permitiendo identificar clústeres, repeticiones y variabilidad real.

- **Línea horizontal de la moda:** Indica el tiempo de respuesta más frecuente. Una moda baja y repetida implica estabilidad; una moda alta indica posibles problemas de rendimiento constante.

  Si existen múltiples modas, se sugiere la coexistencia de distintos comportamientos del sistema.

Conclusión: La moda revela el comportamiento más habitual del backend bajo carga.



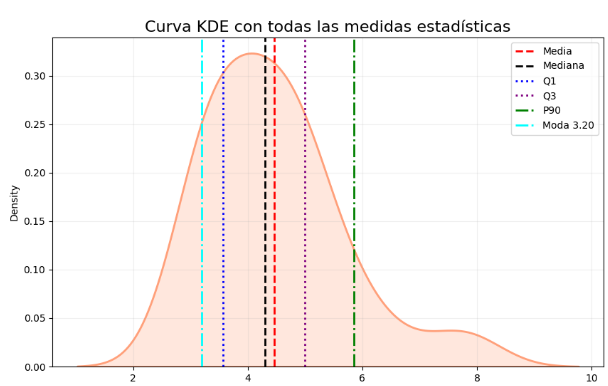
Interpretación del boxplot con Q1, Q2, Q3 y P90

El boxplot resume la distribución mediante:

- **Q1 (Percentil 25%):** indica que el 25% de las peticiones fueron más rápidas.
- **Q2 = Mediana (Percentil 50%):** representa el comportamiento típico.
- **Q3 (Percentil 75%):** valores por debajo de este abarcan el 75% de las peticiones.
- **P90:** el 90% de las peticiones no supera este tiempo. Es crítico para SLA modernos.

Si P90 es mucho mayor que Q3, se evidencia una cola larga, asociada a peticiones muy lentas.

Conclusión: Un boxplot compacto indica estabilidad; una caja amplia o un P90 distante revela inestabilidad y saturación.



Interpretación de la curva KDE con todas las medidas

La KDE actúa como una versión suavizada del histograma:

- Permite observar densidad y tendencias generales.
- Picos representan valores frecuentes; valles representan valores raros.
- Colas largas evidencian problemas de rendimiento.

También facilita comparar:

- Media (sensible a valores altos).
- Mediana (valor central típico).
- Modas múltiples (si existen varios picos).
- Q1, Q3 y P90 en relación con la forma de la distribución.

Conclusión: La KDE permite saber si existe un solo patrón de carga o múltiples comportamientos coexistentes.

Tabla de frecuencias

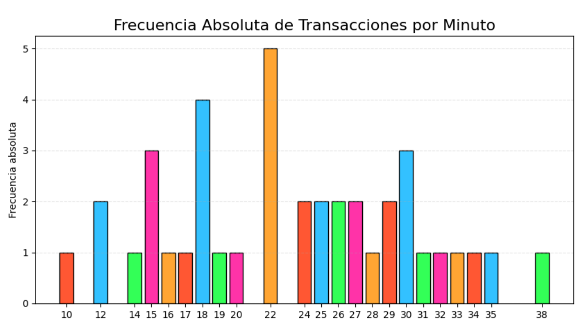
La organización de datos en tablas de frecuencias permite sintetizar valores repetidos y visualizar patrones. Las columnas más típicas son: valor / frecuencia absoluta / frecuencia acumulada / frecuencia relativa / frecuencia relativa acumulada.

Graficos para trabajar Tablas de Frecuencias

Al trabajar con tablas de frecuencias, es fundamental seleccionar una representación gráfica que facilite la interpretación de los datos. A continuación se presentan los gráficos más adecuados y sus principales ventajas.

- **Gráfico de Barras**
  - Es la mejor opción para acompañar una tabla de frecuencias.
  - Permite identificar de manera clara cuáles valores presentan mayor frecuencia.
  - Facilita la comparación visual entre categorías gracias a la diferencia en alturas.
  - Es una representación directa, intuitiva y de rápida comprensión.
  - Ayuda a detectar patrones, picos y valores dominantes dentro del conjunto de datos.
- **Polígono de Frecuencia**
  - Útil para analizar tendencias y la forma de la distribución.
  - Facilita la comparación entre dos o más distribuciones superpuestas.
  - Permite observar si la distribución es creciente, decreciente, simétrica o sesgada.
  - Aporta una visualización más continua que complementa al histograma.

Ejemplo aplicado: Monitoreo de transacciones por minuto en un servicio. Construimos tabla de frecuencias con frecuencia absoluta, acumulada, relativa y relativa acumulada; además gráfico de barras y polígono de frecuencia. Datos: 10, 12, 15, 18, 22, 25, 27, 30, 28, 26, 24, 22, 20, 18, 15, 17, 19, 22, 25, 29, 32, 35, 38, 34, 30, 27, 24, 22, 18, 16, 14, 12, 15, 18, 22, 26, 29, 31, 33, 30



Interpretación del gráfico de barras basado en frecuencia absoluta

El gráfico de barras muestra visualmente la frecuencia absoluta de cada nivel de transacciones. Es la representación más directa de la tabla de frecuencias.

El gráfico revela:

- **Niveles más frecuentes del sistema**

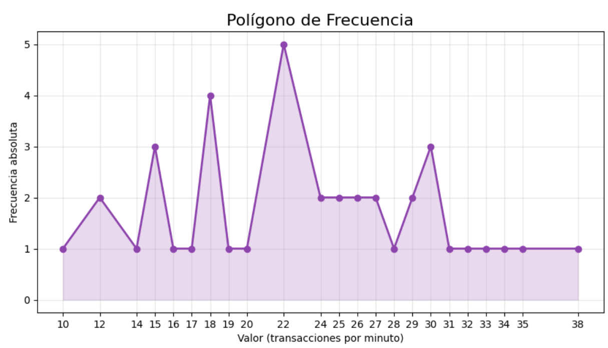
  Las barras más altas indican qué valores aparecen más veces. En el monitoreo de transacciones, esto equivale a los niveles típicos de carga operacional.

- **Identificación de picos operativos**

  Si ciertos valores (por ejemplo, 22, 25 o 30 transacciones/minuto) aparecen repetidamente, el sistema tiende a trabajar regularmente alrededor de esas cargas.

- **Variabilidad de los datos**

  Una distribución con alturas irregulares evidencia fluctuaciones de actividad. Si las barras siguen un patrón estable, el sistema es consistente.



Interpretación del polígono de frecuencias

El polígono de frecuencias conecta los puntos de la frecuencia absoluta formando una curva, ofreciendo una visión más fluida de la distribución.

Permite observar:

- **Tendencias generales**

  La forma del polígono muestra si los niveles de transacción suben progresivamente, caen de forma abrupta, presentan picos aislados o se concentran alrededor de un rango.

- **Estructura del comportamiento del sistema**

  Un polígono con varios picos puede sugerir múltiples estados del sistema: carga normal, carga media o carga alta.

- **Comparación visual entre valores cercanos**

  A diferencia del gráfico de barras, muestra continuidad, útil para identificar patrones ascendentes o descendentes.

Datos agrupados

Cuando existe una gran cantidad de observaciones o valores continuos, los datos se organizan en clases. Esto permite visualizar tendencias globales.

Ejemplo aplicado: Tenemos una estación ambiental con 1000 lecturas de temperatura. Construimos clases (anchura fija), tabla de frecuencias agrupadas, marcas de clase, estimación de media y varianza usando marcas, y graficamos un polígono de frecuencias.

Interpretación de la gráfica:

- El polígono muestra la forma general de la distribución sin necesidad de visualizar los 1000 datos individuales.
- Las marcas de clase se emplean para estimar la media y la varianza de manera aproximada.
- Comparando polígonos entre días, pueden detectarse anomalías en sensores o cambios repentinos en el ambiente.

Visualizaciones para resumen de datos

- **Histogramas:** Muestran la forma de la distribución. Tomamos de ejemplo el registro de los tiempos de las llamadas recibidas en un call center.

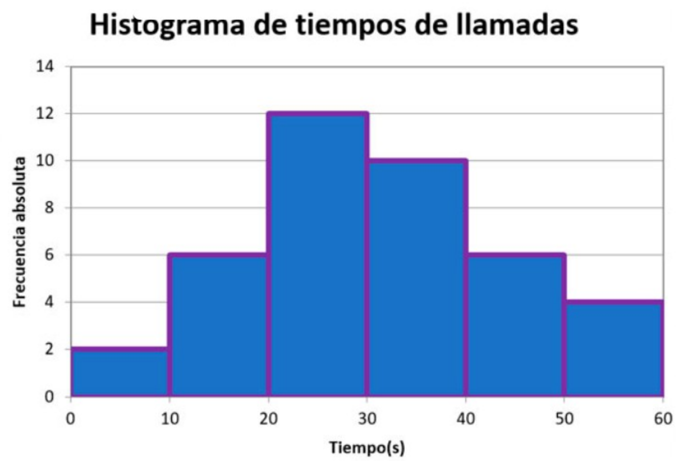


Figure 2: Ejemplo de histograma

- **Polígonos de frecuencia:** Útiles para comparar intervalos o periodos. Tomamos el mismo ejemplo anterior.

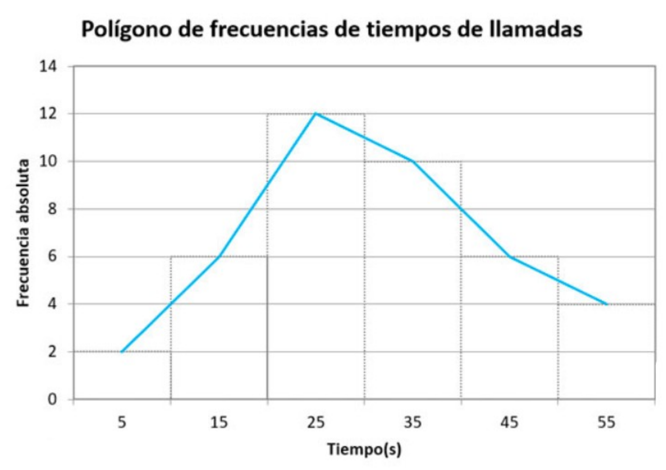


Figure 3: Ejemplo de poligono de frecuencias

- **Boxplot:** Representa mediana, cuartiles, rango intercuartílico y valores atípicos. Usamos las distribución de edad por nivel de educación en una empresa.

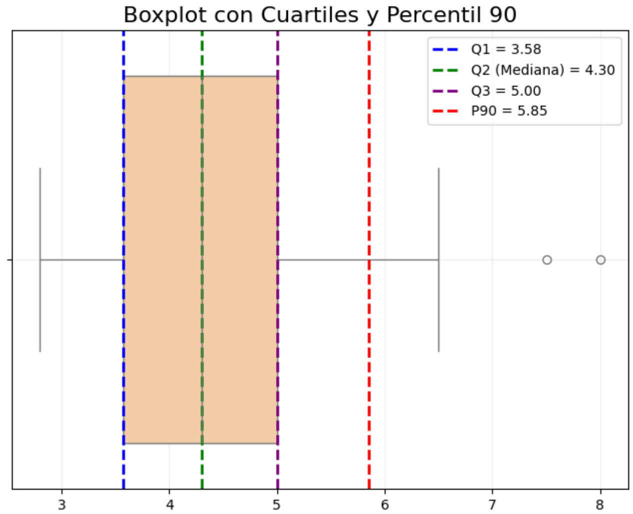


Figure 4: Ejemplo de boxplot

- **Diagramas de tallo y hoja:** Conservan los datos originales y permiten detectar patrones finos. Se toman las edades de personas que se suben a un vagon en especifico.

0	2	1	4	2	3	6					
1	7	8									
2	4	7	0	0	2	6	7	7	1		
3	5	8	6	8	8	1	5	7			
4	2	5	0	0	4	4	8	8	0	0	8
5	7	8	1	8							
6	0	2									
7											
8	4										

Figure 5: Ejemplo de diagrama de tallo y hoja

5. Medidas de tendencia central y dispersión — Ejemplos aplicados

A continuación se presentan ejemplos de la vida real que ilustran cada uno de los conceptos estudiados en esta sección. Cada caso introduce un conjunto de datos realista y explica cómo la medida estadística ayuda a la toma de decisiones en escenarios profesionales.

5.1 Tendencia central

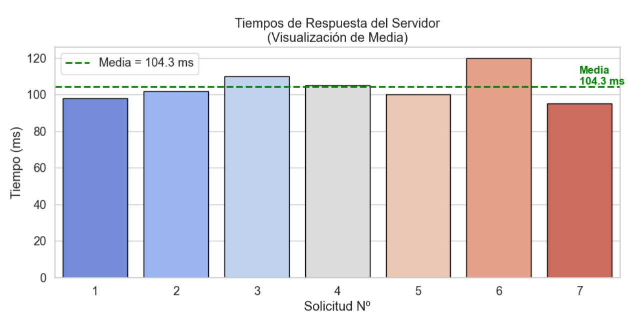
5.1.1 Media — Ejemplo aplicado

Contexto: Una empresa tecnológica registra los tiempos promedio (en milisegundos) que tardan sus servidores en responder a solicitudes críticas durante una prueba de estrés:

$$98,\ 102,\ 110,\ 105,\ 100,\ 120,\ 95.$$

Aplicación real: La media permite estimar el rendimiento general del sistema:

$$\bar{x} = \frac{98 + 102 + 110 + 105 + 100 + 120 + 95}{7} = 104.3\, ms$$



Interpretación: Un tiempo medio de 104.3 ms indica que el servidor opera dentro del margen aceptable (<120 ms). Si la media aumentara, se justificaría una optimización o ampliación de recursos.

5.1.2 Mediana — Ejemplo aplicado

Contexto: El departamento de recursos humanos analiza los salarios mensuales (en millones) de un equipo de 9 desarrolladores:

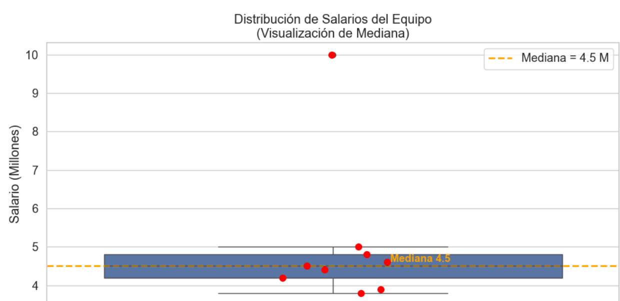
$$3.8,\ 4.2,\ 5.0,\ 4.8,\ 10.0,\ 4.5,\ 3.9,\ 4.4,\ 4.6.$$

Ordenados:

$$3.8,\ 3.9,\ 4.2,\ 4.4,\ 4.5,\ 4.6,\ 4.8,\ 5.0,\ 10.0$$

Aplicación real: La mediana reduce el impacto de salarios atípicos (como 10 millones) y representa mejor el "salario típico" del grupo

$$Mediana = 4.5 millones$$



Interpretación: Aunque existe un salario extremadamente alto, el valor central sugiere que la mayoría se encuentra alrededor de 4.5 millones.

5.1.3 Moda — Ejemplo aplicado

Contexto: Una empresa de comercio electrónico registra cuántas veces se usaron diferentes métodos de pago:

- Tarjeta débito: 125

- Tarjeta crédito: 310
- Nequi: 450
- PSE: 450
- Efectivo: 90

Aplicación real: La moda identifica los métodos más frecuentes:

$$Moda = \{Nequi, PSE\}$$

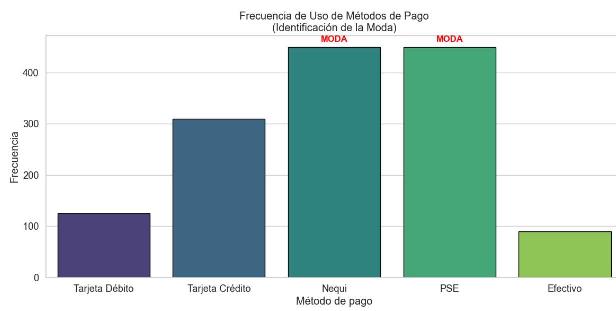


Figure 6: Enter Caption

Interpretación: El sitio debe optimizar especialmente Nequi y PSE, pues son los métodos preferidos.

5.2 Medidas de dispersión

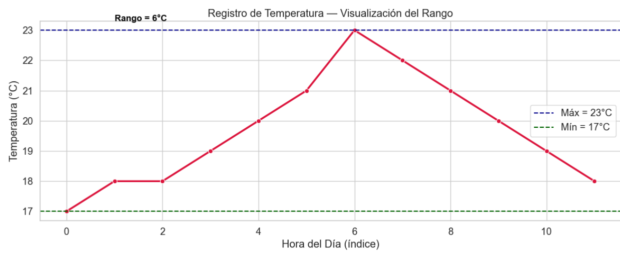
5.2.1 Rango — Ejemplo aplicado

Contexto: Un centro de datos registra la temperatura (°C) durante el día:

$$17,\ 18,\ 18,\ 19,\ 20,\ 21,\ 23,\ 22,\ 21,\ 20,\ 19,\ 18.$$

Aplicación real: El rango permite detectar variaciones peligrosas:

$$R = 23 - 17 = 6°C$$



Interpretación: Aunque la media parezca estable, una variación de 6°C puede indicar fallas de refrigeración.

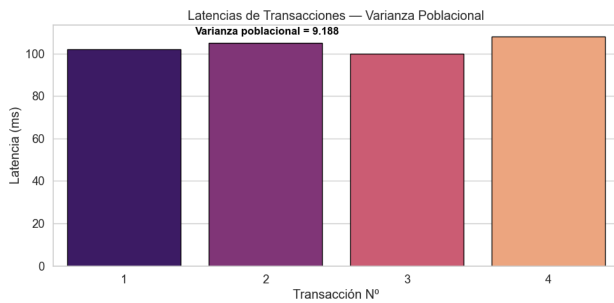
5.2.2 Varianza poblacional — Ejemplo aplicado

Contexto: Una empresa monitorea la latencia exacta de todas las transacciones realizadas en una hora (consideramos que es toda la población). Datos en ms:

$$102,\ 105,\ 100,\ 108.$$

Aplicación real: La varianza poblacional evalúa qué tan uniforme es la experiencia:

$$\sigma^2 = \frac{(102 - 103.75)^2 + (105 - 103.75)^2 + (100 - 103.75)^2 + (108 - 103.75)^2}{4} = 9.1875\, ms^2$$



Interpretación: Una baja varianza indica transacciones estables.

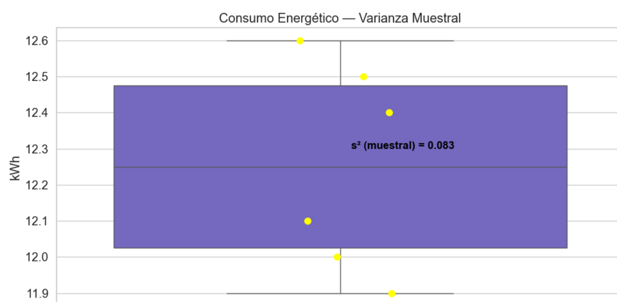
5.2.3 Varianza muestral (estimador insesgado) — Ejemplo aplicado

Contexto: Un ingeniero analiza seis mediciones del consumo energético (kWh):

$$12.1,\ 12.0,\ 12.5,\ 11.9,\ 12.4,\ 12.6.$$

$$\bar{x} = 12.25$$

$$s^2 = \frac{(12.1 - 12.25)^2 + \cdots + (12.6 - 12.25)^2}{5} \approx 0.067$$



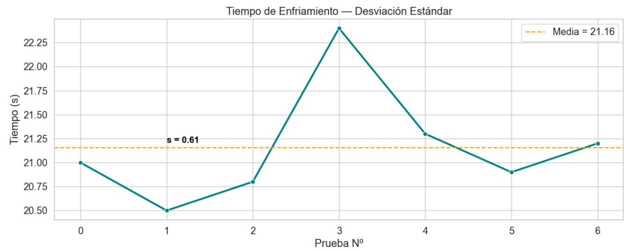
Interpretación: La baja variabilidad es útil para predecir el consumo energético.

5.2.4 Desviación estándar — Ejemplo aplicado

Contexto: Un laboratorio mide el tiempo de enfriamiento de un componente electrónico en 7 pruebas:

$$21.0,\ 20.5,\ 20.8,\ 22.4,\ 21.3,\ 20.9,\ 21.2.$$

$$s = 0.57$$



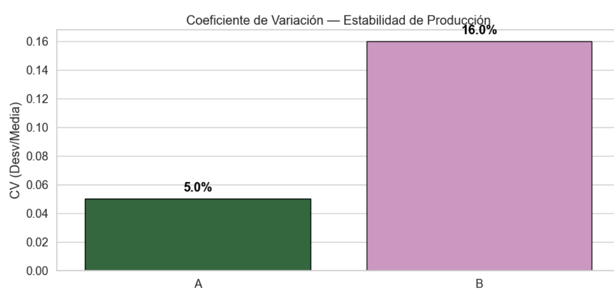
Aplicación real e interpretación: Una desviación estándar tan baja indica alta estabilidad térmica del componente.

5.2.5 Coeficiente de variación (CV) — Ejemplo aplicado

Contexto: Comparación entre dos máquinas que producen tornillos:

Máquina	Media (mm)	Desv. Est. (mm)
A	5.00	0.25
B	5.00	0.80

$$CV_A = \frac{0.25}{5} = 0.05 = 5\% \qquad CV_B = \frac{0.80}{5} = 0.16 = 16\%$$



Interpretación: Aunque ambas máquinas producen con la misma media, la máquina B es mucho más inestable, afectando la calidad.

5.2.6 Desviación absoluta media (MAD) — Ejemplo aplicado

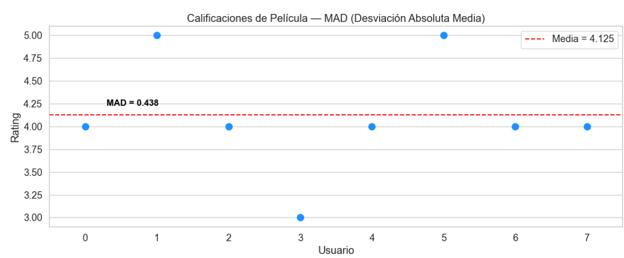
Contexto: Una plataforma de streaming analiza cuánta diferencia hay entre el rating promedio de una película (4.2/5) y las calificaciones de usuarios:

$$4,\ 5,\ 4,\ 3,\ 4,\ 5,\ 4,\ 4.$$

Media:

$$\bar{x} = 4.125$$

$$MAD = \frac{|4 - 4.125| + |5 - 4.125| + \cdots + |4 - 4.125|}{8} = 0.359$$



Interpretación: Un MAD pequeño indica opiniones muy consistentes: la película gusta a la mayoría.

6. Ejemplos de temario agrupados: comparación entre métodos manuales y digitales

Ejemplo 1 — Datos no agrupados (Control de calidad en una línea de producción)

Una empresa de manufactura registra el tiempo de calibración (en segundos) de una máquina CNC. El equipo de Calidad necesita validar si el proceso está estable y si los tiempos cumplen con la ventana operativa del SLA interno. Se toman 20 mediciones no agrupadas en un turno: 12.1, 11.7, 12.4, 13.0, 12.8, 11.9, 12.2, 12.5, 12.7, 12.6, 13.1, 12.9, 11.8, 12.3, 12.4, 12.5, 12.2, 12.1, 12.0, 12.8 Realiza el Procedimiento manual — paso a paso para hallar la media, la mediana, la moda, la varianza, la desviación estándar

A) Procedimiento manual — paso a paso

1. Cálculo de la media

$$\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$$

La suma total es:

$$\sum x_i = 248.0$$

Como $n = 20$:

$$\bar{x} = \frac{248.0}{20} = 12.4$$

2. Cálculo de la mediana

Ordenando los datos de menor a mayor:

$$11.7,\ 11.8,\ 11.9,\ 12.0,\ 12.1,\ 12.1,\ 12.2,\ 12.2,\ 12.3,\ 12.4,$$

$$12.4,\ 12.5,\ 12.5,\ 12.6,\ 12.7,\ 12.8,\ 12.8,\ 12.9,\ 13.0,\ 13.1$$

Para n par, la mediana es:

$$\tilde{x} = \frac{x_{10} + x_{11}}{2} = \frac{12.4 + 12.4}{2} = 12.4$$

3. Cálculo de la moda

Se identifican las frecuencias:

$$12.1,\ 12.2,\ 12.4,\ 12.5,\ 12.8$$

cada una aparece dos veces.

$$Moda = \{12.1,\ 12.2,\ 12.4,\ 12.5,\ 12.8\}$$

4. Cálculo de varianza y desviación estándar

La tabla de desviaciones respecto a la media ($\bar{x} = 12.4$):

$$\sum (x_i - \bar{x})^2 = 3.10$$

Varianza poblacional

$$\sigma^2 = \frac{1}{n} \sum (x_i - \bar{x})^2$$

$$\sigma^2 = \frac{3.10}{20} = 0.155$$

Varianza muestral

$$s^2 = \frac{1}{n-1} \sum (x_i - \bar{x})^2$$

$$s^2 = \frac{3.10}{19} = 0.16315789$$

Desviación estándar

$$\sigma = \sqrt{0.155} = 0.3937$$

$$s = \sqrt{0.16315789} = 0.4039$$

Resultados finales

$$Media = 12.4$$

$$Mediana = 12.4$$

$$Moda = \{12.1,\ 12.2,\ 12.4,\ 12.5,\ 12.8\}$$

$$\sigma^2 = 0.155, \qquad \sigma = 0.3937$$

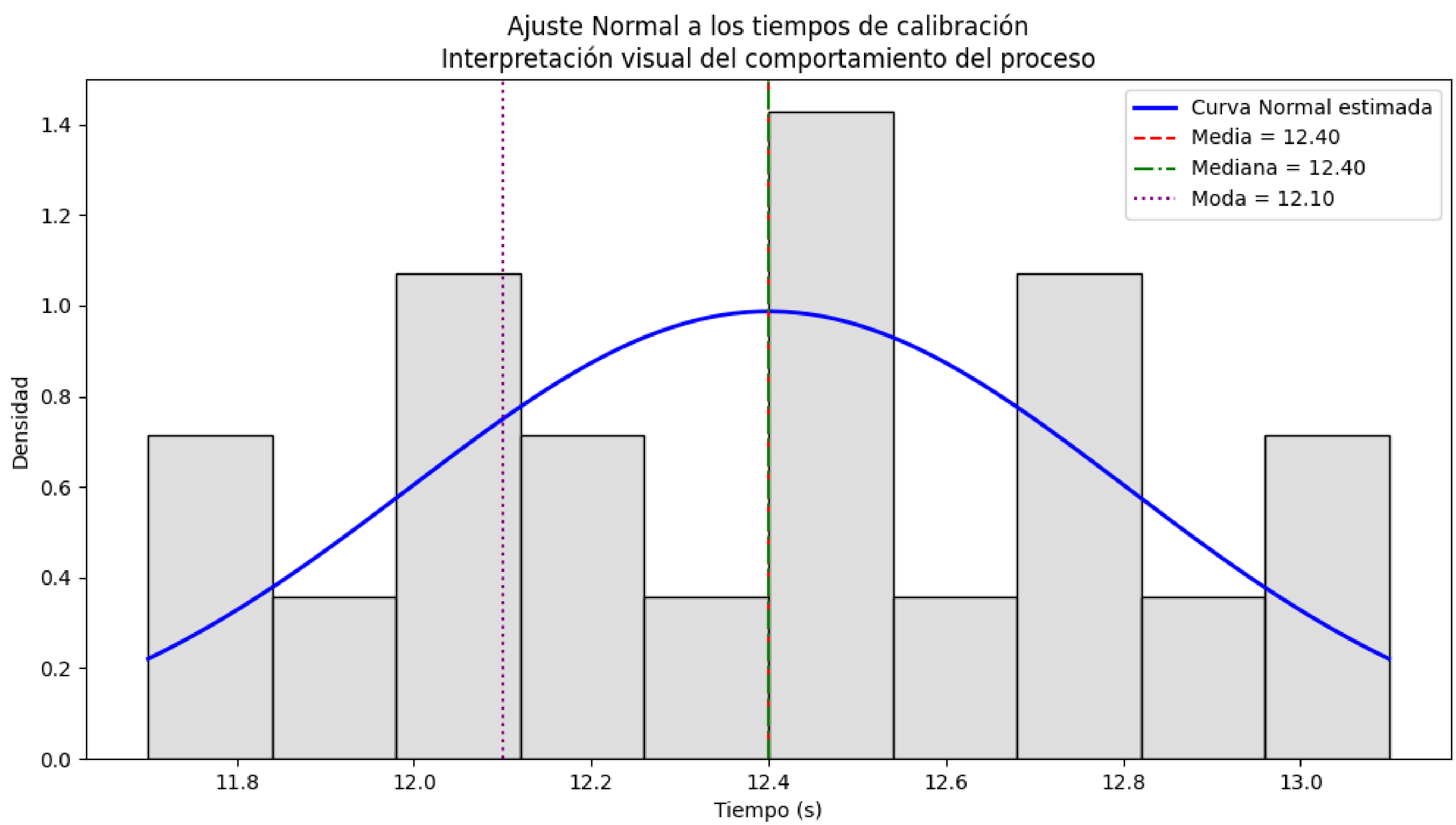
$$s^2 = 0.16316, \qquad s = 0.4039$$

A continuación se resuelve el ejemplo con Python

```python
import numpy as np                  # 1: numpy — arrays y utilidades numéricas
import pandas as pd                 # 2: pandas — estructuras tabulares (DataFrame)
import seaborn as sns               # 3: seaborn — visualización estadística de alto nivel
import matplotlib.pyplot as plt     # 4: matplotlib — backend de gráficos
from scipy.stats import norm, mode # 5: norm para la PDF Normal; mode para la moda

# 1) Datos: 20 mediciones no agrupadas
data = [                           # 6: lista Python con las 20 observaciones (tiempos en s)
    12.1, 11.7, 12.4, 13.0, 12.8, 11.9, 12.2, 12.5,
    12.7, 12.6, 13.1, 12.9, 11.8, 12.3, 12.4, 12.5,
    12.2, 12.1, 12.0, 12.8
]

df = pd.DataFrame({"tiempo": data})  # 7: convierto la lista en DataFrame con columna "tiempo"

# 2) Cálculos estadísticos
media = df["tiempo"].mean()          # 8: media aritmética (E[X]) — indicador central
mediana = df["tiempo"].median()      # 9: mediana (50%) — robusta frente a outliers

# scipy.stats.mode devuelve un array, tomamos solo el valor
moda_valor = mode(df["tiempo"], keepdims=True).mode[0]  # 10: moda (valor más frecuente)

var_poblacional = df["tiempo"].var(ddof=0)  # 11: varianza poblacional (n en el denominador)
var_muestral = df["tiempo"].var(ddof=1)     # 12: varianza muestral (n-1 en denom — estimador insesgado)

desv_poblacional = df["tiempo"].std(ddof=0) # 13: desviación estándar poblacional
desv_muestral = df["tiempo"].std(ddof=1)    # 14: desviación estándar muestral

# 15-21: impresión de resultados clave en consola para control operativo
print("Media:", media)
print("Mediana:", mediana)
print("Moda:", moda_valor)
print("Varianza poblacional:", var_poblacional)
print("Varianza muestral:", var_muestral)
print("Desv. estándar poblacional:", desv_poblacional)
print("Desv. estándar muestral:", desv_muestral)

#  Gráfico: Ajuste tipo Normal (curva teórica)
plt.figure(figsize=(12, 6))              # 22: tamaño del lienzo — formato apaisado para presentación

# Histograma base
sns.histplot(df["tiempo"], kde=False, color="lightgray", stat="density", bins=10)  # 23:
# - histograma de densidad: muestra la distribución empírica
# - stat="density": normaliza área a 1 para superponer PDF teórica

# Rango de valores
x = np.linspace(min(data), max(data), 200)  # 24: vector x continuo entre min y max con 200 puntos

# Curva normal con parámetros estimados
pdf = norm.pdf(x, media, desv_muestral)     # 25: calculo la PDF normal usando media y desviación muestral
plt.plot(x, pdf, color="blue", linewidth=2, label="Curva Normal estimada")  # 26: dibujo la curva

# Líneas de referencia estadísticas
plt.axvline(media, color="red", linestyle="--", label=f"Media = {media:.2f}")     # 27: media
plt.axvline(mediana, color="green", linestyle="-.", label=f"Mediana = {mediana:.2f}") # 28: mediana
plt.axvline(moda_valor, color="purple", linestyle=":", label=f"Moda = {moda_valor:.2f}") # 29: moda

plt.title("Ajuste Normal a los tiempos de calibración\nInterpretación visual del comportamiento del proceso") # 30
plt.xlabel("Tiempo (s)")                 # 31: etiqueta eje X
plt.ylabel("Densidad")                   # 32: etiqueta eje Y (densidad porque usamos stat='density')
plt.legend()                             # 33: leyenda con líneas de referencia
plt.show()                               # 34: renderizo gráfica en pantalla
```

A continuación, se presentan los resultados obtenidos a partir del análisis estadístico de los datos en Python, Asimismo, se muestra la gráfica correspondiente que permiten visualizar el comportamiento general del conjunto de valores.

```
Media: 12.4
Mediana: 12.4
Moda: 12.1
Varianza poblacional: 0.15500000000000005
Varianza muestral: 0.16315789473684217
Desv. estándar poblacional: 0.3937003937005906
Desv. estándar muestral: 0.40392808114420836
```



Ejemplo 2 — Datos agrupados (marcas y frecuencias)

Con el objetivo de evaluar la estabilidad operativa del proceso de *picking* en un centro de distribución durante una semana de alta demanda, se registraron 50 observaciones de los tiempos de ciclo (en minutos) asociados a la operación. Los datos fueron organizados en intervalos homogéneos para facilitar el análisis estadístico y determinar el nivel de variabilidad del proceso. La siguiente tabla consolida los intervalos, sus marcas de clase y las frecuencias observadas:

$Intervalo(min)$	$Marca(x_i)$	$Frecuencia(f_i)$
$6-8$	7	12
$8-10$	9	18
$10-12$	11	14
$12-14$	13	6
Total	$-$	**50**

A partir de la distribución anterior, se solicita calcular:

- La media agrupada.
- La varianza muestral.
- La desviación estándar muestral.

Procedimiento Manual

1. Cálculo de la media agrupada

La media agrupada se calcula mediante:

$$\bar{x} = \frac{\sum f_i x_i}{\sum f_i}$$

Primero se obtiene la suma ponderada:

$$\sum f_i x_i = (12)(7) + (18)(9) + (14)(11) + (6)(13)$$

$$\sum f_i x_i = 84 + 162 + 154 + 78 = 478$$

El total de observaciones es:

$$N = \sum f_i = 50$$

Por tanto:

$$\bar{x} = \frac{478}{50} = 9.56$$

2. Cálculo de la varianza muestral para datos agrupados

La fórmula utilizada es:

$$s^2 = \frac{\sum f_i (x_i - \bar{x})^2}{N - 1}$$

Se calcula cada desviación al cuadrado ponderada:

$$(x_1 - \bar{x})^2 = (7 - 9.56)^2 = 6.5536$$

$$(x_2 - \bar{x})^2 = (9 - 9.56)^2 = 0.3136$$

$$(x_3 - \bar{x})^2 = (11 - 9.56)^2 = 2.0736$$

$$(x_4 - \bar{x})^2 = (13 - 9.56)^2 = 11.8176$$

Multiplicadas por sus frecuencias:

$$\sum f_i (x_i - \bar{x})^2 = (12)(6.5536) + (18)(0.3136) + (14)(2.0736) + (6)(11.8176)$$

$$\sum f_i(x_i - \bar{x})^2 = 78.6432 + 5.6448 + 29.0304 + 70.9056$$

$$\sum f_i(x_i - \bar{x})^2 = 184.224$$

Ahora se divide entre $N - 1$:

$$s^2 = \frac{184.224}{49} = 3.7597$$

3. Cálculo de la desviación estándar

$$s = \sqrt{s^2} = \sqrt{3.7597} = 1.939$$

Resultados Finales

$$\bar{x} = 9.56$$

$$s^2 = 3.7597$$

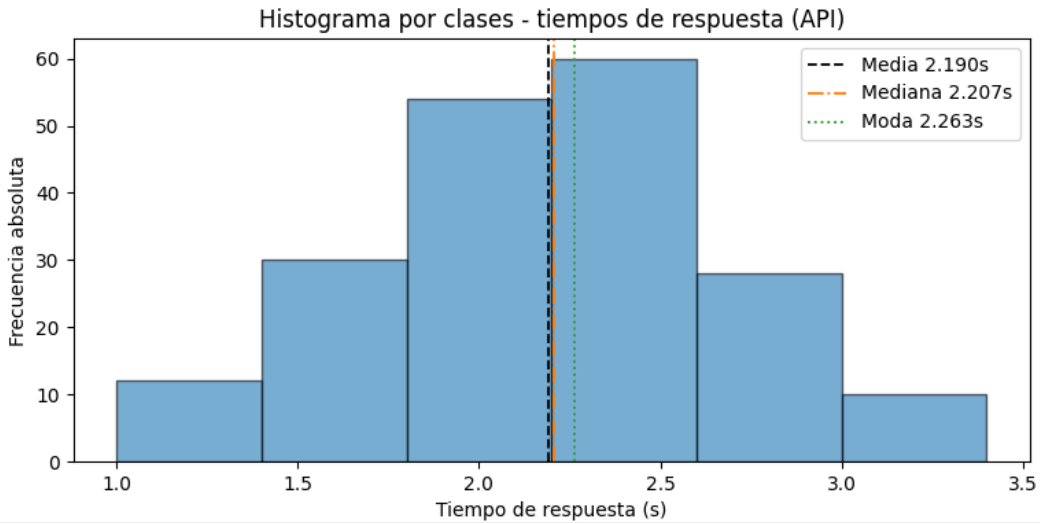
$$s = 1.939$$

Los indicadores muestran una dispersión moderada del proceso respecto al rango objetivo de 8 a 10 minutos, lo que sugiere revisar la estandarización de métodos en turnos con mayor variabilidad.

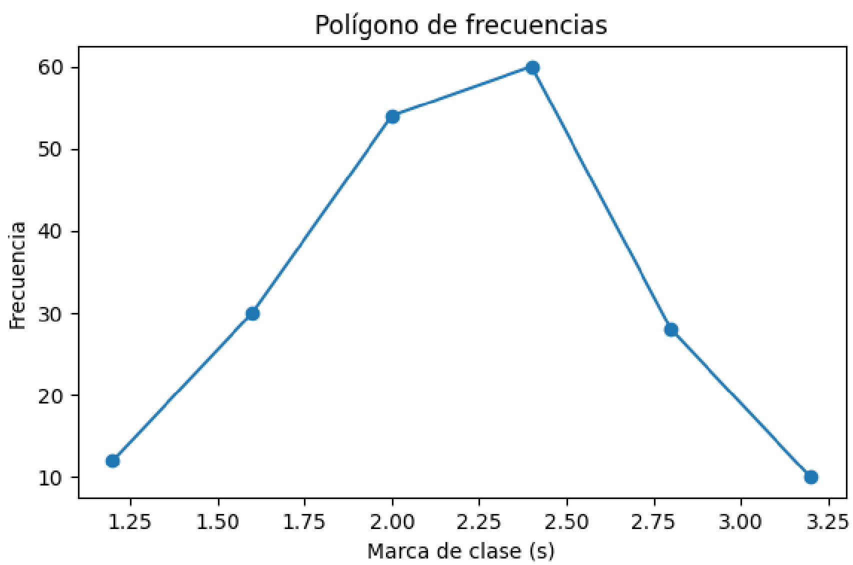
A continuación se resuelve el ejemplo con Python

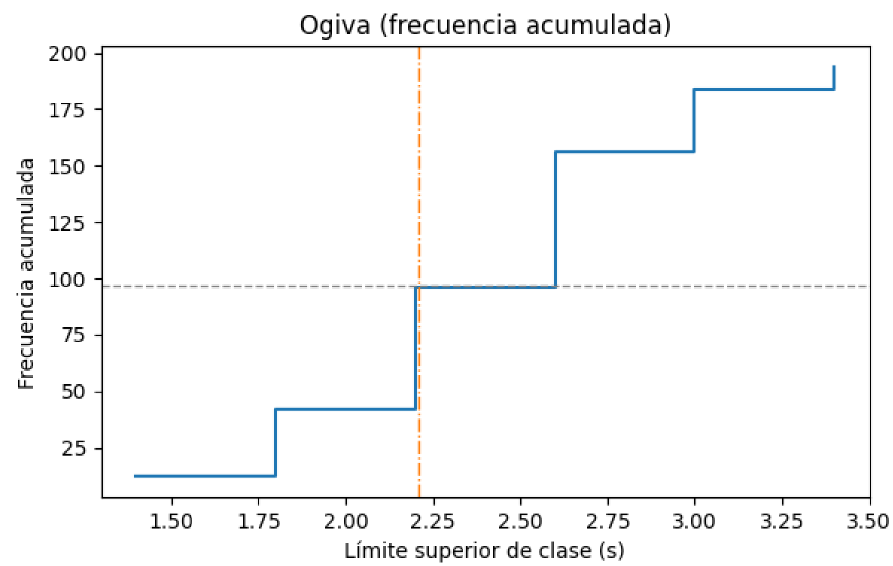
```python
import numpy as np                 # operaciones numéricas eficientes (arrays, sqrt, cumsum, linspace)
import matplotlib.pyplot as plt    # generación de gráficos

# --- Entrada: datos agrupados por clases ---
intervalos = [(1.0,1.4),(1.4,1.8),(1.8,2.2),(2.2,2.6),(2.6,3.0),(3.0,3.4)]
marcas = np.array([1.2,1.6,2.0,2.4,2.8,3.2])  # marca de clase (punto medio de cada intervalo)
f = np.array([12,30,54,60,28,10])             # frecuencias absolutas por clase
N = f.sum()                                   # tamaño total de la muestra (suma de frecuencias)
h = intervalos[0][1] - intervalos[0][0]       # ancho de clase (se asume constante)

# --- Estadísticos agrupados (media, varianza, desviación) ---
mean_agr = (marcas * f).sum() / N             # media por marcas: sum(x_i * f_i) / N
var_num = (f * (marcas - mean_agr)**2).sum()  # numerador varianza agrupada
var_muestral = var_num / (N - 1)              # varianza muestral (n-1): estimador insesgado
std_muestral = np.sqrt(var_muestral)          # desviación estándar muestral

# --- Mediana por interpolación en clase ---
pos_med = N / 2                               # posición de la mediana en frecuencia acumulada
cumf = np.cumsum(f)                           # frecuencia acumulada por clases
med_class_idx = np.searchsorted(cumf, pos_med) # índice de la clase mediana (primera cumf >= N/2)
L_med = intervalos[med_class_idx][0]          # límite inferior de la clase mediana
F_b = cumf[med_class_idx-1] if med_class_idx > 0 else 0  # frecuencia acumulada previa (F_{b})
f_m = f[med_class_idx]                        # frecuencia de la clase mediana
median_interp = L_med + ((pos_med - F_b) / f_m) * h  # interpolación lineal dentro de la clase

# --- Moda por interpolación (fórmula triangular) ---
modal_idx = np.argmax(f)                      # índice de la clase modal (máxima frecuencia)
L_mod = intervalos[modal_idx][0]              # límite inferior de la clase modal
f_prev = f[modal_idx-1] if modal_idx-1 >= 0 else 0   # frecuencia de la clase anterior (f_{i-1})
f_next = f[modal_idx+1] if modal_idx+1 < len(f) else 0 # frecuencia de la clase siguiente (f_{i+1})
mode_interp = L_mod + ((f[modal_idx] - f_prev) / (2*f[modal_idx] - f_prev - f_next)) * h

# --- Coeficiente de variación y sesgo (Pearson) ---
cv = std_muestral / mean_agr                  # coeficiente de variación (relativo)
skew_pearson = 3 * (mean_agr - median_interp) / std_muestral  # fórmula de Pearson (asimetría aproximada)

# --- Resumen impreso ---
print(f"N = {N}")
print(f"Media agrupada = {mean_agr:.6f} s")
print(f"Varianza muestral = {var_muestral:.6f} s^2")
print(f"Desviación estándar = {std_muestral:.6f} s")
print(f"Mediana interpolada = {median_interp:.6f} s")
print(f"Moda interpolada = {mode_interp:.6f} s")
print(f"CV = {cv:.4f} ({cv*100:.1f}%)")
print(f"Skewness (Pearson) = {skew_pearson:.3f}")

# --- Gráficos ---
# 1) Histograma por clases (barras por frecuencia) + marcas y líneas centrales
fig, ax = plt.subplots(figsize=(9,4))
left_edges = [iv[0] for iv in intervalos]    # límites inferiores para ubicar barras
right_edges = [iv[1] for iv in intervalos]   # límites superiores (no usados directamente en bar)
widths = [r-l for l,r in intervalos]         # anchos de cada clase

# barras (frecuencia absoluta)
ax.bar(left_edges, f, width=widths, align='edge', alpha=0.6, edgecolor='black')
ax.set_xlabel('Tiempo de respuesta (s)')
ax.set_ylabel('Frecuencia absoluta')
ax.set_title('Histograma por clases - tiempos de respuesta (API)')

# anotar media, mediana, moda
ax.axvline(mean_agr, color='k', linestyle='--', linewidth=1.3, label=f"Media {mean_agr:.3f}s")
ax.axvline(median_interp, color='C1', linestyle='-.', linewidth=1.3, label=f"Mediana {median_interp:.3f}s")
ax.axvline(mode_interp, color='C2', linestyle=':', linewidth=1.3, label=f"Moda {mode_interp:.3f}s")
ax.legend()

# 2) Polígono de frecuencias y ogiva en figura aparte
fig2, (ax1, ax2) = plt.subplots(1,2, figsize=(12,4))

# Polígono: puntos en marcas conectados
ax1.plot(marcas, f, marker='o', linestyle='-', linewidth=1.5)
ax1.set_xlabel('Marca de clase (s)')
ax1.set_ylabel('Frecuencia')
ax1.set_title('Polígono de frecuencias')

# Ogiva (frecuencia acumulada) — step en límites superiores
cum_rel = np.cumsum(f)
ax2.step([iv[1] for iv in intervalos], cum_rel, where='post', linewidth=1.5)
ax2.set_xlabel('Límite superior de clase (s)')
ax2.set_ylabel('Frecuencia acumulada')
ax2.set_title('Ogiva (frecuencia acumulada)')
ax2.axhline(N/2, color='gray', linestyle='--', linewidth=1)  # referencia horizontal N/2
ax2.axvline(median_interp, color='C1', linestyle='-.', linewidth=1)  # marca la mediana interpolada

plt.tight_layout()
plt.show()
```

A continuación, se presentan los resultados obtenidos a partir del análisis estadístico de los datos en Python, Asimismo, se muestra la gráfica correspondiente que permiten visualizar el comportamiento general del conjunto de valores.




```
N = 194
Media agrupada = 2.189691 s
Varianza muestral = 0.245696 s^2
Desviación estándar = 0.495678 s
Mediana interpolada = 2.206667 s
Moda interpolada = 2.263158 s
CV = 0.2264 (22.6%)
Skewness (Pearson) = -0.103
```





Ejemplo 3 — Percentiles, cuartiles y boxplot (Salarios)

Se analiza la estructura salarial de una pequeña organización compuesta por tres equipos. El objetivo es caracterizar la posición relativa de los salarios mediante percentiles (Q1, Q2, Q3, P90), medir la dispersión con el rango intercuartílico (IQR) y detectar posibles valores atípicos usando la regla de Tukey. Las conclusiones orientarán decisiones de compensación y auditoría de nómina.

Datos (salarios en millones) – tamaño de la muestra combinado $n = 36$:

$$\begin{aligned} EquipoA(12): &\quad 3.2,\ 3.3,\ 3.5,\ 3.6,\ 3.8,\ 4.0,\ 4.1,\ 4.2,\ 4.4,\ 4.5,\ 4.8,\ 5.0 \\ EquipoB(12): &\quad 4.5,\ 4.6,\ 4.7,\ 4.8,\ 5.0,\ 5.2,\ 5.3,\ 5.5,\ 5.7,\ 6.0,\ 6.5,\ 7.5 \\ EquipoC(12): &\quad 2.8,\ 3.0,\ 3.1,\ 3.2,\ 3.3,\ 3.4,\ 3.6,\ 3.8,\ 4.0,\ 4.1,\ 4.9,\ 8.0 \end{aligned}$$

Para facilitar el cálculo manual se usa la lista ordenada (posiciones 1 a 36):

$$\begin{aligned} 1: &\quad 2.8,\ 2: 3.0,\ 3: 3.1,\ 4: 3.2,\ 5: 3.2,\ 6: 3.3,\ 7: 3.3,\ 8: 3.4, \\ 9: &\quad 3.5,\ 10: 3.6,\ 11: 3.6,\ 12: 3.8,\ 13: 3.8,\ 14: 4.0,\ 15: 4.0, \\ 16: &\quad 4.1,\ 17: 4.1,\ 18: 4.2,\ 19: 4.4,\ 20: 4.5,\ 21: 4.5,\ 22: 4.6, \\ 23: &\quad 4.7,\ 24: 4.8,\ 25: 4.8,\ 26: 4.9,\ 27: 5.0,\ 28: 5.0,\ 29: 5.2, \\ 30: &\quad 5.3,\ 31: 5.5,\ 32: 5.7,\ 33: 6.0,\ 34: 6.5,\ 35: 7.5,\ 36: 8.0. \end{aligned}$$

Convención utilizada para posiciones: Para un percentil de orden p ($0 < p < 1$) se usa la posición:

$$k = p \cdot (n + 1).$$

Si k no es entero, se interpola linealmente entre las observaciones adyacentes.

Procedimiento manual — Cálculos paso a paso

1. Primer cuartil Q_1 (p = 0.25)

$$k_1 = 0.25 \cdot (36 + 1) = 0.25 \cdot 37 = 9.25.$$

La posición 9.25 está entre la observación en posición 9 y la 10:

$$x_{(9)} = 3.5, \quad x_{(10)} = 3.6.$$

Interpolación:

$$Q_1 = x_{(9)} + (0.25)(x_{(10)} - x_{(9)}) = 3.5 + 0.25(3.6 - 3.5) = 3.5 + 0.025 = 3.525.$$

2. Mediana Q_2 (p = 0.50)

$$k_2 = 0.50 \cdot 37 = 18.5.$$

Entre la posición 18 y 19:

$$x_{(18)} = 4.2, \quad x_{(19)} = 4.4.$$

Interpolación:

$$Q_2 = 4.2 + 0.5(4.4 - 4.2) = 4.2 + 0.1 = 4.30.$$

3. Tercer cuartil Q_3 (p = 0.75)

$$k_3 = 0.75 \cdot 37 = 27.75.$$

Entre la posición 27 y 28:

$$x_{(27)} = 5.0, \quad x_{(28)} = 5.0.$$

Como ambos valores son iguales:

$$Q_3 = 5.00.$$

4. Rango intercuartílico (IQR)

$$\mathrm{IQR} = Q_3 - Q_1 = 5.00 - 3.525 = 1.475.$$

5. Percentil 90 (P90)

$$k_{90} = 0.90 \cdot 37 = 33.3.$$

Entre la posición 33 y 34:

$$x_{(33)} = 6.0, \quad x_{(34)} = 6.5.$$

Interpolación:

$$P_{90} = 6.0 + 0.3(6.5 - 6.0) = 6.0 + 0.15 = 6.15.$$

6. Detección de outliers — regla de Tukey (1.5 $\times$ IQR)

$$Lmiteinferior = Q_1 - 1.5 \cdot \mathrm{IQR} = 3.525 - 1.5(1.475) = 1.3125.$$

$$Lmitesuperior = Q_3 + 1.5 \cdot \mathrm{IQR} = 5.00 + 1.5(1.475) = 7.2125.$$

Observaciones fuera del intervalo $[1.3125,\ 7.2125]$ se consideran atípicas. En la muestra:

$$7.5 > 7.2125, \quad 8.0 > 7.2125,$$

por tanto $7.5 y 8.0$ son **outliers** superiores.

Resultados finales (resumen)

$$Q_1 = 3.525, Q_2\,(Mediana) = 4.30, Q_3 = 5.00, \mathrm{IQR} = 1.475, P_{90} = 6.15, Outliers(Tukey) = \{7.5,\ 8.0\}.$$

Interpretación ejecutiva (breve)

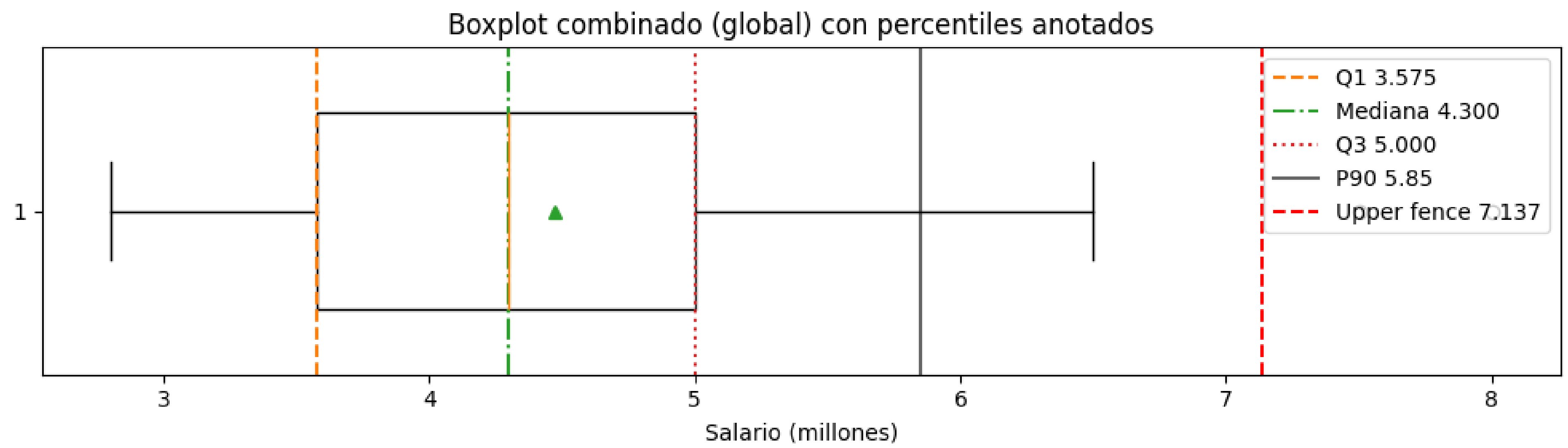
La mitad central del personal gana entre ≈ 3.53 y 5.00 millones; la mediana (4.30 M) es el punto de referencia central. El IQR indica dispersión moderada en la mitad central. Se detectan dos salarios atípicos por encima del límite superior (7.5 y 8.0 M), lo que sugiere auditar su origen (bonos, contratos o errores) antes de incorporarlos en decisiones de política salarial o cálculo de medias.

A continuación se resuelve el ejemplo con Python

```python
# ejemplo3_percentiles_boxplot.py
import numpy as np                          # numpy: operaciones numéricas vectorizadas
import pandas as pd                         # pandas: estructura DataFrame para resumen
import matplotlib.pyplot as plt             # matplotlib: visualización

# --- Datos ---
team_A = np.array([3.2,3.3,3.5,3.6,3.8,4.0,4.1,4.2,4.4,4.5,4.8,5.0])   # Equipo A
team_B = np.array([4.5,4.6,4.7,4.8,5.0,5.2,5.3,5.5,5.7,6.0,6.5,7.5])   # Equipo B
team_C = np.array([2.8,3.0,3.1,3.2,3.3,3.4,3.6,3.8,4.0,4.1,4.9,8.0])   # Equipo C

# combinado
all_data = np.concatenate([team_A, team_B, team_C])  # concatena los 3 equipos en un vector
n = len(all_data)                                    # tamaño total de la muestra

# --- Cálculo NumPy (método digital) ---
# Nota: en numpy moderno usar `method='linear'` en lugar de `interpolation`; aquí se mantiene por claridad.
q1_np = np.percentile(all_data, 25, interpolation='linear')   # primer cuartil (25%)
q2_np = np.percentile(all_data, 50, interpolation='linear')   # mediana (50%)
q3_np = np.percentile(all_data, 75, interpolation='linear')   # tercer cuartil (75%)
p90_np = np.percentile(all_data, 90, interpolation='linear')  # percentil 90
iqr_np = q3_np - q1_np                                        # IQR

# fences Tukey (regla 1.5*IQR para outliers)
lower_fence = q1_np - 1.5 * iqr_np
upper_fence = q3_np + 1.5 * iqr_np

# identificar outliers (valores fuera de fences)
outliers = all_data[(all_data < lower_fence) | (all_data > upper_fence)]

# Resumen en DataFrame para reporte
summary = pd.DataFrame({
    'N': [n],
    'Q1': [q1_np],
    'Q2_median': [q2_np],
    'Q3': [q3_np],
    'IQR': [iqr_np],
    'P90': [p90_np],
    'LowerFence': [lower_fence],
    'UpperFence': [upper_fence],
    'Outliers': [outliers.tolist()]
})

print(summary.T)   # transpuesto para lectura vertical en consola

# --- Boxplots comparativos por equipo ---
fig, ax = plt.subplots(figsize=(9,5))
ax.boxplot([team_A, team_B, team_C],
           labels=['Equipo A','Equipo B','Equipo C'],
           showmeans=True)   # showmeans dibuja un marcador para la media
ax.set_title('Comparativo Boxplot por Equipo (salarios, millones)')
ax.set_ylabel('Salario (millones)')

# Anotar outliers detectados (global)
# Atención: la x fija 3.05 está pensada para colocar la anotación cerca del tercer boxplot (ajustable)
for val in outliers:
    ax.annotate(f'Outlier {val:.2f}', xy=(3.05, val), xytext=(3.25, val+0.2),
                arrowprops=dict(arrowstyle='->', lw=0.8), fontsize=9)

plt.grid(axis='y', linestyle=':', alpha=0.5)
plt.tight_layout()
plt.show()

# --- Boxplot combinado con líneas de percentiles globales ---
fig2, ax2 = plt.subplots(figsize=(10,3))
ax2.boxplot(all_data, vert=False, showmeans=True, widths=0.6)  # boxplot horizontal
ax2.set_xlabel('Salario (millones)')
ax2.set_title('Boxplot combinado (global) con percentiles anotados')

# líneas verticales (percentiles y fence superior)
ax2.axvline(q1_np, color='C1', linestyle='--', label=f'Q1 {q1_np:.3f}')
ax2.axvline(q2_np, color='C2', linestyle='-.', label=f'Mediana {q2_np:.3f}')
ax2.axvline(q3_np, color='C3', linestyle=':', label=f'Q3 {q3_np:.3f}')
ax2.axvline(p90_np, color='k', linestyle='-', alpha=0.6, label=f'P90 {p90_np:.2f}')
ax2.axvline(upper_fence, color='r', linestyle='--', label=f'Upper fence {upper_fence:.3f}')
ax2.legend(loc='upper right')

plt.tight_layout()
plt.show()
```

A continuación, se presentan los resultados obtenidos a partir del análisis estadístico de los datos en Python, Asimismo, se muestra la gráfica correspondiente que permiten visualizar el comportamiento general del conjunto de valores.

```
N                36
Q1            3.575
Q2_median       4.3
Q3              5.0
IQR           1.425
P90            5.85
LowerFence   1.4375
UpperFence   7.1375
Outliers   [7.5, 8.0]
```



7. Conclusiones

La estadística descriptiva constituye la base fundamental del análisis de datos, ya que permite transformar información cruda en conocimiento comprensible mediante la recolección, organización, resumen y presentación de los datos. A lo largo del tiempo, ha evolucionado desde los antiguos censos y registros estatales hasta convertirse en una herramienta científica indispensable en todas las disciplinas, formalizándose gracias a los aportes de autores como Quetelet, Pearson y Galton, quienes consolidaron los conceptos de tendencia central, dispersión y correlación. En el ámbito de la ingeniería de sistemas, la estadística descriptiva es esencial para la gestión eficiente de datos y el diseño de soluciones informáticas, ya que permite analizar el rendimiento de los sistemas, detectar problemas y establecer métricas de desempeño. Los avances recientes integran esta rama con la inteligencia artificial, la automatización y el análisis de grandes volúmenes de datos (Big Data), mediante herramientas capaces de generar visualizaciones interactivas y análisis automáticos. En la actualidad, la estadística descriptiva cumple también una función ética y comunicativa al fomentar decisiones basadas en evidencia y en la interpretación responsable de la información. En síntesis, sigue siendo una herramienta científica vigente y en expansión, cuyo desarrollo continúa de la mano de la ciencia de datos y la inteligencia artificial, fortaleciendo las competencias analíticas necesarias para el ejercicio profesional de la ingeniería moderna.

8. Referencias

- Akal, E. (2020, octubre 20). *DS-STAR: A state-of-the-art, versatile data science agent*. Recuperado de https://www.nocierreslosojos.com/estadistica-moderna-historia-galton-pearson/
- Flourish. (s. f.). *Flourish*. Recuperado de https://flourish.studio/
- IBM. (2025). *¿Qué es el análisis exploratorio de datos?*. Recuperado de https://www.ibm.com/es-es/think/topics/exploratory-data-analysis
- López, J. F. (2019, noviembre 15). *Origen de la estadística—Qué es y su impacto histórico*. Economipedia. Recuperado de https://economipedia.com/definiciones/origen-estadistica.html
- Pareja, C., & Sevilla Arias, A. (2025, mayo 29). *Estadística descriptiva: Qué es, tipos y ejemplos*. Economipedia. Recuperado de https://economipedia.com/definiciones/estadistica-descriptiva.html
- Tableau. (2021). *¿Qué es Tableau?*. Recuperado de https://www.tableau.com/es-es/why-tableau/what-is-tableau
- Triola, M. F. (2018). *Estadística* (12.ª ed.). Pearson.
- Yoon, J., & Nam, J. (2025, noviembre 6). *DS-STAR: A state-of-the-art versatile data science agent*. Google Cloud. Recuperado de https://research.google/blog/ds-star-a-state-of-the-art-versatile-data-science-agent/