

Notas de clase Business Analytics

Daniel Parra

Invalid Date

Tabla de contenidos

Bienvenidos

Este libro ha sido creado con el propósito de ofrecer un resumen claro y conciso de los conceptos clave de las clases de Business Analytics para los posgrados de administración de la Pontificia Universidad Javeriana. Es un complemento diseñado para reforzar el aprendizaje en clase, pero no pretende reemplazar la experiencia educativa que estas ofrecen.

Parte del texto en estas notas de clase ha sido elaborado con ayuda de inteligencia artificial (ChatGPT), combinando apuntes propios y referencias de diversos libros. Las ilustraciones, inspiradas en el libro “Math with Bad Drawings” de Ben Orlin, también han sido generadas utilizando inteligencia artificial.

© 2025 Pontificia Universidad Javeriana. Este libro está licenciado bajo una licencia CC BY 4.0.

1 Introducción: Descubriendo la Analítica de Datos

Imagina que trabajas en una empresa que debe decidir qué producto lanzar o qué precio poner a sus artículos. ¿Cómo tomas estas decisiones? Antiguamente, las decisiones se tomaban según la intuición del jefe (**HiPPO**, la persona mejor pagada en la organización), pero actualmente contamos con algo mejor: la analítica de datos.

1.1. ¿Qué es la Analítica de Datos para Negocios?

La analítica de datos es la aplicación de herramientas tecnológicas y estadísticas para analizar información relevante y apoyar la toma de decisiones en una organización.

Los datos, sin embargo, no son suficientes por sí solos. Necesitamos interpretarlos de manera efectiva para contar historias convincentes que permitan tomar decisiones informadas.

1.2. ¿Qué tipo de preguntas podemos responder con la analítica?

La analítica de negocios nos permite responder preguntas prácticas, como:

- ¿Quiénes son mis clientes?
- ¿Qué campaña de marketing es más efectiva?
- ¿Cuál sería el efecto de cambiar los precios?
- ¿Cómo predecir la demanda de un producto?

1.3. Tipos de Analítica

Hay tres tipos principales de analítica:

1.3.1. Analítica Descriptiva

Describe la situación actual mediante visualizaciones y estadísticas básicas.

Ejemplos:

- ¿Han crecido las ventas este año?
- ¿Qué región vende menos?

1.3.2. Analítica Predictiva

Predice qué sucederá, aprovechando correlaciones entre variables. Aquí no es necesaria la causalidad, sino una relación estadística sólida.

Ejemplos:

- ¿Cuál es la probabilidad de que un cliente deje de comprar?
- ¿Qué tan probable es que una transacción sea fraudulenta?

1.3.3. Analítica Prescriptiva

Sugiere decisiones óptimas, enfocándose en relaciones causales y, frecuentemente, utilizando experimentos.

Ejemplos:

- ¿Subir el precio incrementará las ganancias?
- ¿Qué diseño de página web aumentará las ventas?

1.4. La cadena de valor en Business Analytics

La analítica de negocios incluye:

- **Recolección y organización de datos:** Crear bases de datos (SQL, Power BI).
- **Análisis de datos:** Usar herramientas como R, Excel o Python para análisis estratégicos.
- **Comunicación de resultados:** Presentar hallazgos en reportes y visualizaciones (PowerPoint, Tableau, Quarto).

2 Fundamentos de Variables y Visualización de Datos

2.1. Introducción a Variables y Datos

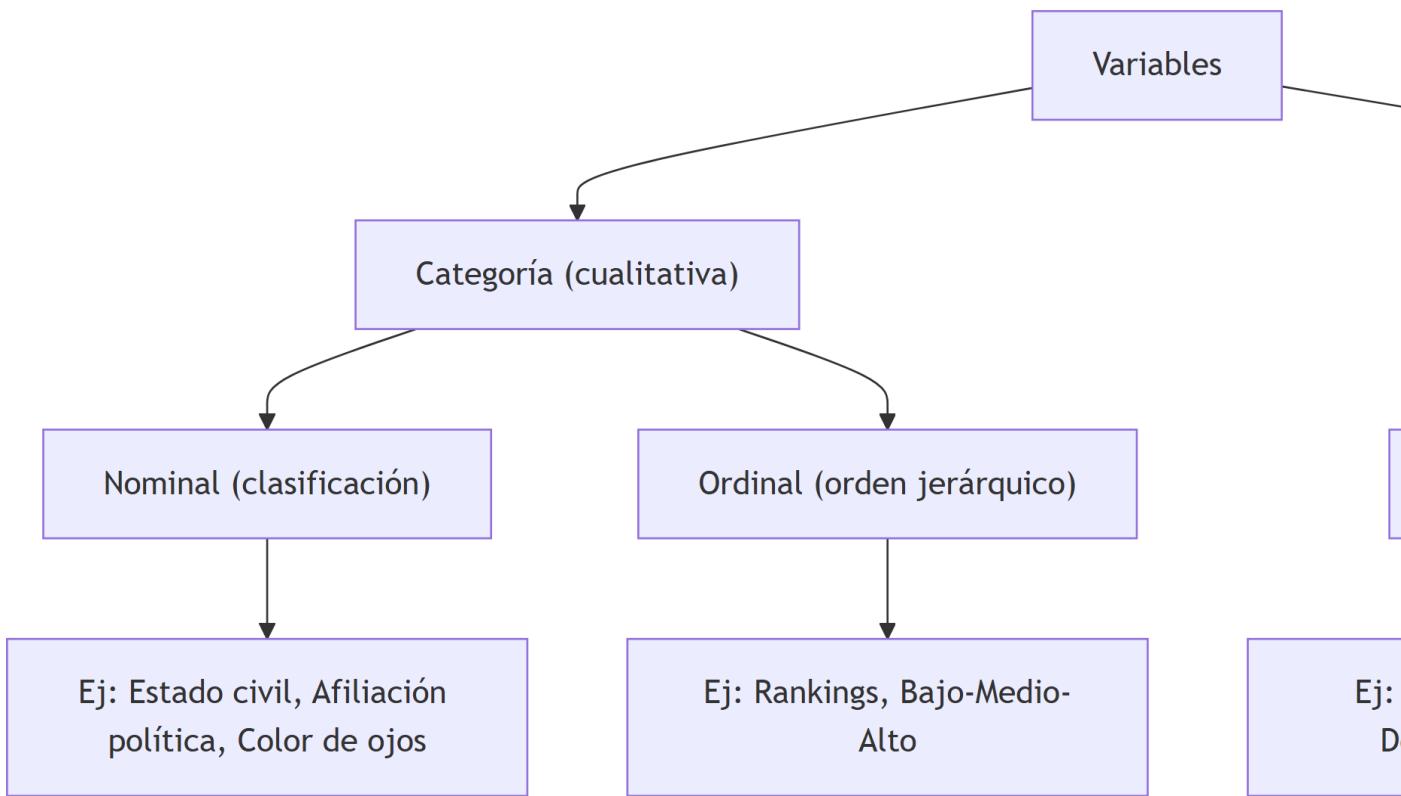
Antes de manipular o graficar, es importante entender qué son los datos y las variables.

Definiciones clave

- **Variable:** Característica que puede cambiar entre observaciones (ej. género, ventas, PIB).
- **Datos:** Conjunto de observaciones (filas) con valores para esas variables (columnas).

2.1.1. Tipos de variables

Las variables se clasifican en dos grandes categorías según su naturaleza:



Clasificación de Variables

Entender el tipo de variable es clave porque:

- **Define el resumen estadístico adecuado:** No calculamos una media de colores de ojos.
- **Determina el gráfico apropiado:** Un histograma para variables numéricas, un gráfico de barras para categorías.
- **Evita errores de interpretación:** Un ordinal tratado como numérico puede dar conclusiones equivocadas.

Ejemplos de tipos de variables:

Tabla 2.1: Ejemplo: Tipos de Variables en una Encuesta

| Pregunta | Respuesta | Tipo de Variable |
|--|---|-----------------------|
| ¿Tiene perfil de Facebook? | Sí / No | Categórica Nominal |
| ¿Cuántos mensajes de texto ha enviado en los últimos dos días? | _____ (número entero) | Numérica Discreta |
| ¿Cuánto tiempo le tomó bajar la aplicación? | _____ (minutos o segundos) | Numérica Continua |
| ¿Cómo evaluaría su experiencia en Facebook? | Muy mala, Mala, Regular, Buena, Muy buena | Categórica Ordinal |

2.2. Organización de Datos Categóricos

2.2.1. Tablas para Datos Categóricos

Tabla resumen: Muestra frecuencias o porcentajes para cada categoría de la variable. A continuación mostramos cómo se distribuyen las calificaciones de nuestro restaurante en Google Maps, basadas en 100 reseñas:

Tabla 2.2: Resumen de Calificaciones en Google Maps

| Calificación | Frecuencia | % sobre 100 reseñas |
|--------------|------------|---------------------|
| Excelente | 40 | 40 % |
| Buena | 35 | 35 % |
| Regular | 20 | 20 % |
| Mala | 5 | 5 % |

Tabla de contingencia: Relaciona dos o más variables categóricas. Muestra la frecuencia con la que ocurren combinaciones de categorías. Por ejemplo, las calificaciones según la hora del día:

Tabla 2.3: Calificaciones por Hora del Día

| Calificación | Mañana (n, %) | Tarde (n, %) | Noche (n, %) | Madrugada (n, %) |
|--------------|---------------|--------------|--------------|------------------|
| Excelente | 10 (40 %) | 15 (50 %) | 10 (33.3 %) | 5 (62.5 %) |
| Buena | 8 (32 %) | 10 (33.3 %) | 8 (26.7 %) | 2 (25 %) |
| Regular | 4 (16 %) | 3 (10 %) | 7 (23.3 %) | 1 (12.5 %) |

| Calificación | Mañana (n, %) | Tarde (n, %) | Noche (n, %) | Madrugada (n, %) |
|--------------|-------------------|-------------------|-------------------|------------------|
| Mala | 3 (12 %) | 2 (6.7 %) | 5 (16.7 %) | 0 (0 %) |
| Total | 25 (100 %) | 30 (100 %) | 30 (100 %) | 8 (100 %) |

Al analizar la tabla de contingencia, podemos comparar cómo varían las calificaciones según la hora del día. Por ejemplo, la madrugada muestra un 62.5 % de reseñas “Excelente” y ninguna “Mala”, lo que sugiere que quienes califican en ese horario son clientes muy satisfechos (posiblemente menos volumen y mejor atención). En cambio, durante la noche la proporción de calificaciones “Excelente” baja a 33.3 % y aumentan las “Mala” (16.7 %), lo que podría indicar saturación operativa o tiempos de espera más largos.

La tarde destaca por concentrar el mayor número de reseñas totales y un 50 % de valoraciones “Excelente”, lo que la convierte en una franja estratégica para mantener estándares altos. Estos patrones permiten priorizar recursos: reforzar personal nocturno y replicar buenas prácticas de la madrugada y la tarde.

2.2.2. Visualizando datos categóricos

2.2.2.1. Gráfico de Barras

Los gráficos de barras son la herramienta más versátil para visualizar variables categóricas. Utilizan barras rectangulares cuya longitud (o altura) es proporcional a la frecuencia o porcentaje de cada categoría. Su principal ventaja es la facilidad para comparar visualmente diferentes categorías, ya que el ojo humano puede detectar rápidamente diferencias en longitud.

Las barras pueden orientarse vertical u horizontalmente, y son especialmente útiles cuando los nombres de las categorías son largos (mejor horizontal) o cuando queremos enfatizar el orden jerárquico de los valores (mejor vertical). A diferencia de otros gráficos, las barras deben comenzar siempre desde cero para no distorsionar las comparaciones visuales. Mira un ejemplo de gráfico de barras:

2.2.2.2. Gráfico de Torta (Pie Chart)

Los gráficos de torta representan datos categóricos como sectores de un círculo, donde cada sector es proporcional al porcentaje que representa esa categoría del total. La “torta” completa suma siempre 100 %, y cada “rebanada” muestra visualmente qué proporción ocupa cada categoría.

Su principal fortaleza es mostrar la **composición del total** - es decir, cómo se distribuye un todo entre sus partes. Son especialmente útiles cuando queremos enfatizar que una categoría domina sobre las demás o cuando la pregunta clave es “¿qué porcentaje del total representa

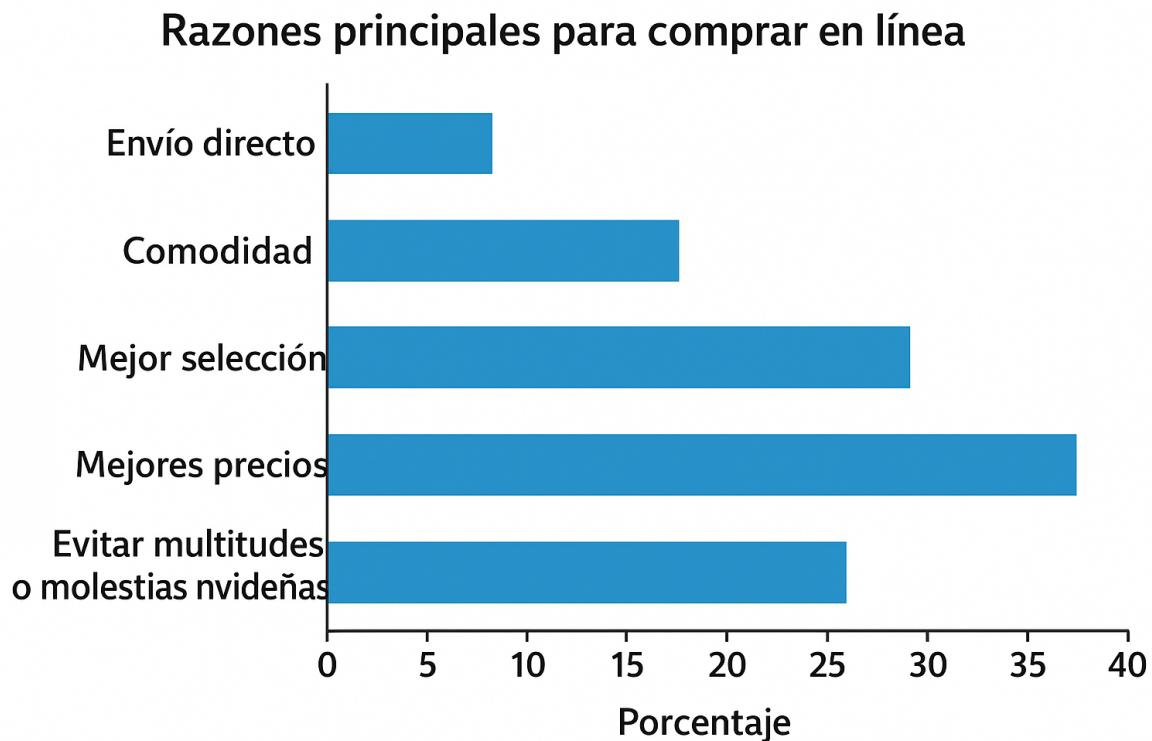


Figura 2.1: Ejemplo de gráfico de barras mostrando la distribución de calificaciones

cada grupo?”. Por ejemplo, para mostrar la participación de mercado de diferentes marcas o la distribución del presupuesto entre departamentos. Este sería el gráfico de torta que muestra la misma información del gráfico de barras anterior:

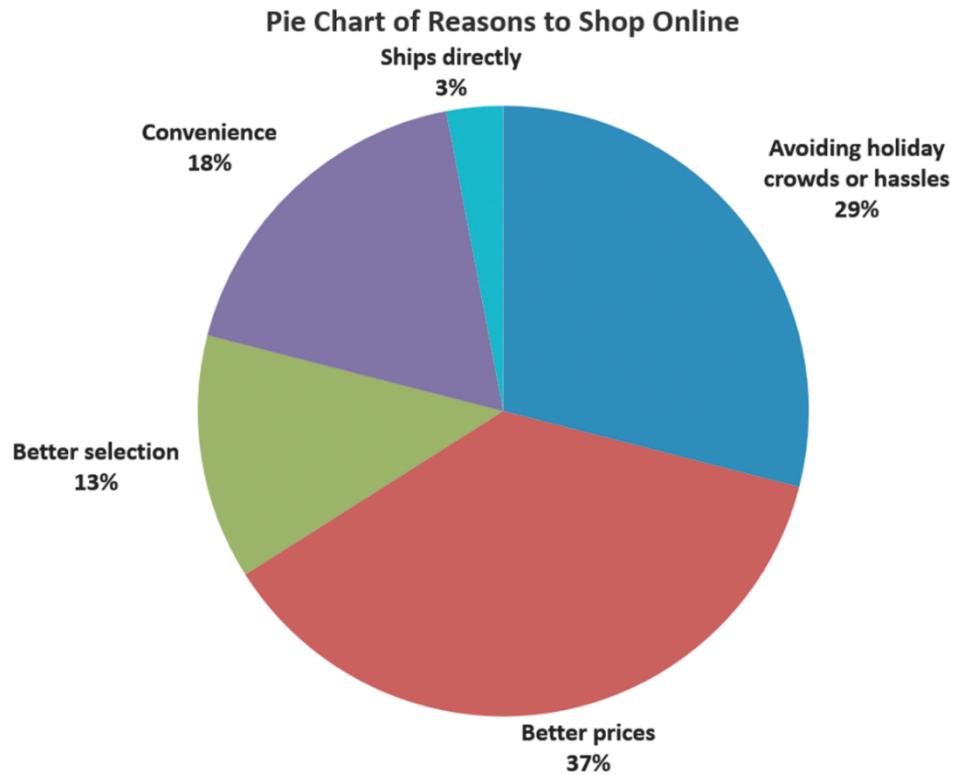


Figura 2.2: Ejemplo de gráfico de torta mostrando la distribución de calificaciones

⚠ Limitaciones del gráfico de torta

- **Difícil comparación de sectores similares:** El ojo humano no distingue bien diferencias pequeñas entre ángulos
- **Máximo recomendado:** No más de 5-7 categorías para mantener legibilidad
- **Mejor alternativa:** Considera un gráfico de barras si necesitas comparar categorías con valores muy cercanos

Los gráficos de torta funcionan mejor cuando hay una o dos categorías claramente dominantes y el resto son minoritarias. Si todas las categorías tienen tamaños similares, un gráfico de barras será más efectivo para la comparación visual.

2.2.2.3. Gráfico de Pareto

El gráfico de Pareto combina barras y una línea para identificar las categorías más importantes según el **principio de Pareto** (regla 80-20). Las barras muestran las frecuencias ordenadas de mayor a menor, mientras que la línea representa el porcentaje acumulado hasta llegar al 100 %.

Su objetivo principal es **priorizar** - ayuda a identificar las pocas categorías que concentran la mayor parte del problema o fenómeno. Por ejemplo, en control de calidad, unas pocas causas suelen generar la mayoría de los defectos; en ventas, unos pocos productos pueden generar la mayor parte de los ingresos.

Elementos clave del gráfico de Pareto:

- **Barras ordenadas:** De mayor a menor frecuencia (izquierda a derecha)
- **Línea acumulada:** Muestra el porcentaje que suman las categorías hasta ese punto
- **Principio 80-20:** Busca el punto donde pocas categorías explican la mayoría del fenómeno

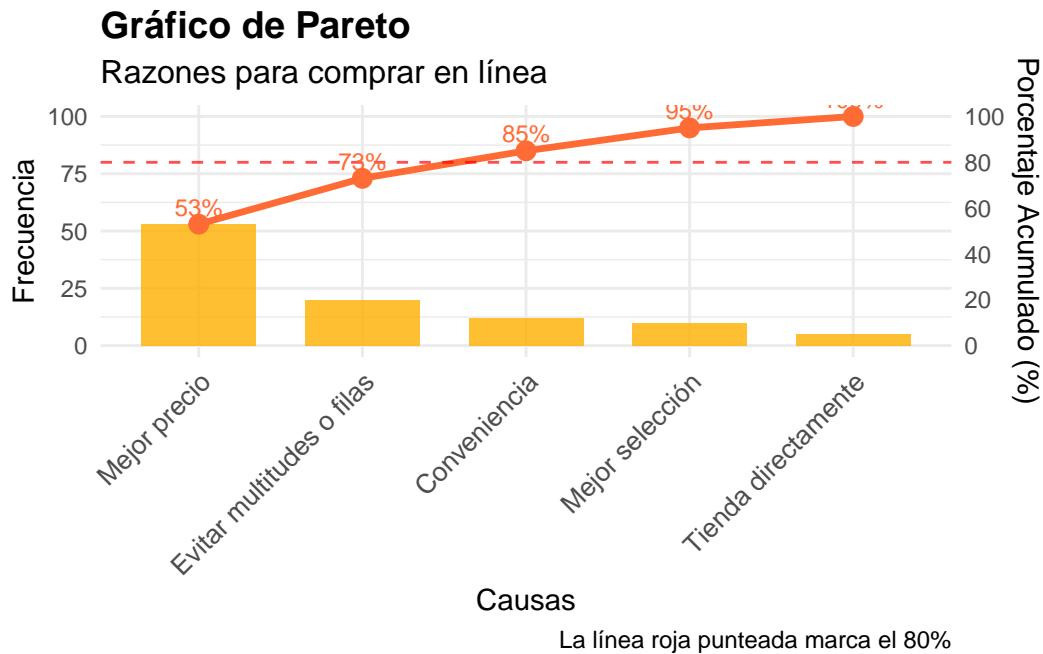


Figura 2.3: Gráfico de Pareto - Análisis de Causas de Defectos

💡 ¿Cuándo usar un gráfico de Pareto?

- **Identificar prioridades:** ¿Qué problemas atender primero?
- **Análisis de causas:** ¿Qué factores tienen mayor impacto?
- **Optimización de recursos:** ¿Dónde enfocar los esfuerzos?
- **Seguimiento de mejoras:** ¿Las acciones redujeron las causas principales?

Ejemplo práctico: Si analizamos las quejas de un restaurante y encontramos que “comida fría” y “tiempo de espera” representan el 75 % de todas las quejas, sabemos que resolver estos dos problemas tendrá el mayor impacto en la satisfacción del cliente. El gráfico de Pareto haría visible esta concentración de manera inmediata.

2.2.3. Visualizando datos categóricos: Gráfico de Barras Emparejadas

El **gráfico de barras emparejadas** (o agrupadas) se utiliza cuando queremos comparar cómo se distribuye una variable categórica principal dentro de los niveles de otra variable categórica. Cada grupo del eje horizontal representa una categoría de la primera variable y, dentro de cada grupo, las barras muestran las categorías de la segunda.

Ejemplo: Si analizamos las calificaciones de un restaurante (Excelente, Buena, Regular, Mala) según la hora del día en que se dejó la reseña (Mañana, Tarde, Noche), podemos ver rápidamente en qué horario se concentran más las opiniones positivas.

Este tipo de gráfico ayuda a detectar patrones de comparación entre grupos: diferencias claras en alturas de barras indican posibles áreas de mejora o fortalezas.

Cuándo usarlo:

- Cuando tienes **dos variables categóricas**.
- Para comparar proporciones o frecuencias entre subgrupos.
- Como complemento visual de una tabla de contingencia.

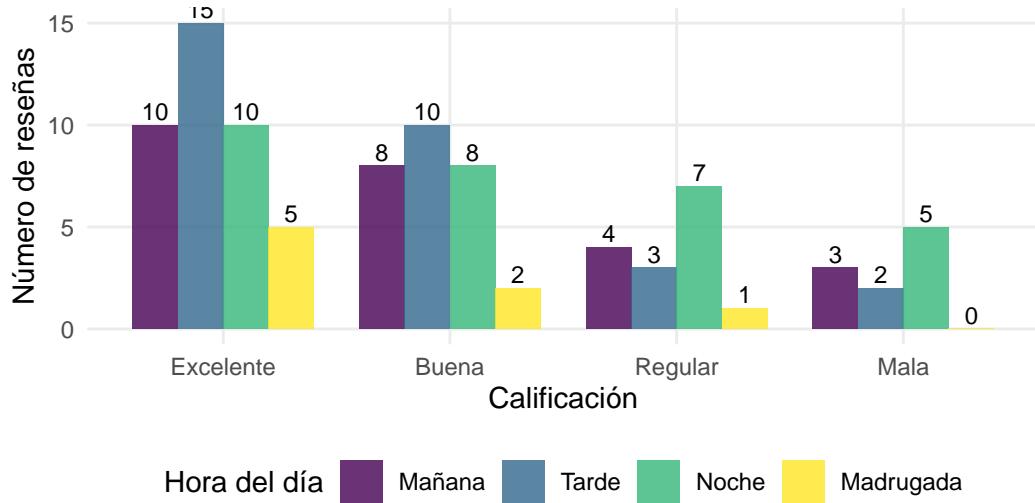
Recomendaciones:

- Ordena las categorías de forma lógica (por tiempo, intensidad, etc.).
- Usa una leyenda clara.
- Evita el exceso de colores; prioriza el contraste.

En resumen: antes de elegir cómo mostrar tus datos categóricos, conviene pensar cuántas variables estás analizando. El siguiente esquema resume las opciones más comunes: si trabajas con una sola variable, puedes resumirla con una tabla y luego elegir un gráfico de barras, de torta o de Pareto según el objetivo. Si tienes dos variables categóricas, una tabla de contingencia es el punto de partida y suele representarse con un gráfico de barras emparejadas para comparar grupos.

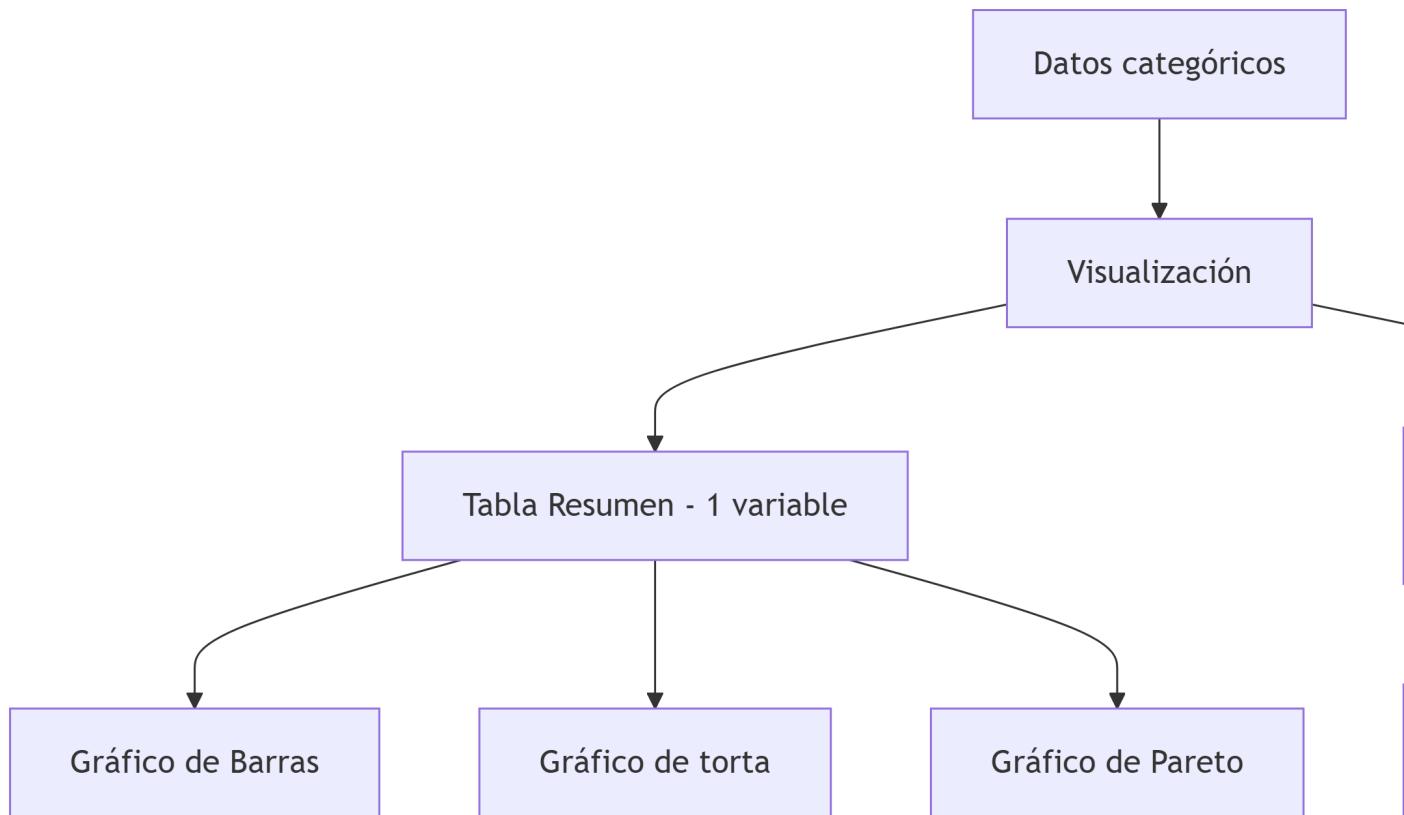
Calificaciones del Restaurante por Hora del Día

Distribución de reseñas según el momento de evaluación



Los datos muestran patrones claros: la madrugada tiene más 'Excelentes' proporcionalmente

Figura 2.4: Gráfico de Barras Emparejadas - Calificaciones por Hora del Día



Opciones de Visualización para Datos Categóricos

2.2.4. Visualizando datos numéricos

2.2.4.1. Distribución de frecuencia

Imagina que registras el **monto gastado por cada cliente** en tu restaurante durante un día. Tienes muchos valores sueltos y quieres resumirlos para entender mejor los patrones.

La **distribución de frecuencia** es una tabla que agrupa esos montos en **categorías numéricas ordenadas** (también llamadas clases) y cuenta cuántos clientes caen en cada una.

¿Cómo elegir las categorías?

Primero necesitas definir **categorías adecuadas**: decidir sus límites (fronteras) y su **ancho**. El número de categorías depende del **rango** de los datos (máximo – mínimo). Si los gastos van desde \$5 hasta \$55, el rango es 50.

Cuando el rango es grande solemos usar más categorías; en la práctica, entre **5 y 15** funciona bien.

¿Cómo calcular el ancho?

Si decides usar, por ejemplo, 10 categorías:

$$\text{Ancho} = \frac{\text{Rango}}{\text{Número de categorías}} = \frac{50}{10} = 5$$

Eso significa que tus clases podrían ser: \$5 - <\$10, \$10 - <\$15, \$15 - <\$20, ... hasta \$55. Luego cuentas cuántos clientes gastaron en cada intervalo y obtienes una tabla clara para analizar tendencias (por ejemplo, “la mayoría gasta entre \$15 y \$25”).

Esta tabla te permite pasar del desorden de datos individuales a una visión estructurada que facilita la toma de decisiones.

Ejemplo práctico: Distribución de gastos

Supongamos que registramos el gasto de **100 clientes** y los agrupamos usando intervalos de \$5:

Tabla 2.4: Distribución de Gastos por Cliente

| Intervalo (USD) | Frecuencia | % Relativo | Frec. Acumulada | % Acumulado |
|-----------------|------------|------------|-----------------|-------------|
| \$5 - <\$10 | 8 | 8 % | 8 | 8 % |
| \$10 - <\$15 | 15 | 15 % | 23 | 23 % |
| \$15 - <\$20 | 22 | 22 % | 45 | 45 % |
| \$20 - <\$25 | 25 | 25 % | 70 | 70 % |

| Intervalo (USD) | Frecuencia | % Relativo | Frec. Acumulada | % Acumulado |
|-----------------|------------|--------------|-----------------|-------------|
| \$25 - <\$30 | 14 | 14 % | 84 | 84 % |
| \$30 - <\$35 | 9 | 9 % | 93 | 93 % |
| \$35 - <\$40 | 5 | 5 % | 98 | 98 % |
| \$40 - <\$45 | 2 | 2 % | 100 | 100 % |
| Total | 100 | 100 % | - | - |

Interpretación clave:

- **Concentración:** El 70 % de los clientes gastó menos de \$25
- **Cola:** Solo el 7 % gastó más de \$35
- **Patrón:** La mayoría de clientes (67 %) gasta entre \$10-\$30

2.2.4.2. Histograma

Siguiendo con el ejemplo de los gastos en el restaurante, el **histograma** es la versión gráfica de la tabla de distribución de frecuencia.

En lugar de mostrar los números en una tabla, dibujamos una barra para cada categoría (intervalo de gasto).

- El **eje horizontal** muestra los intervalos: \$5-<\$10, \$10-<\$15, etc.
- El **eje vertical** muestra la frecuencia (o el porcentaje) de clientes en cada intervalo.
- Las barras van **pegadas** porque los intervalos son continuos: representan rangos de una misma variable numérica.

Un **histograma** es la representación gráfica de una distribución de frecuencias para datos numéricos continuos. A diferencia del gráfico de barras (para datos categóricos), el histograma muestra barras adyacentes sin espacios entre ellas, lo que refleja la naturaleza continua de los datos.

Características clave del histograma:

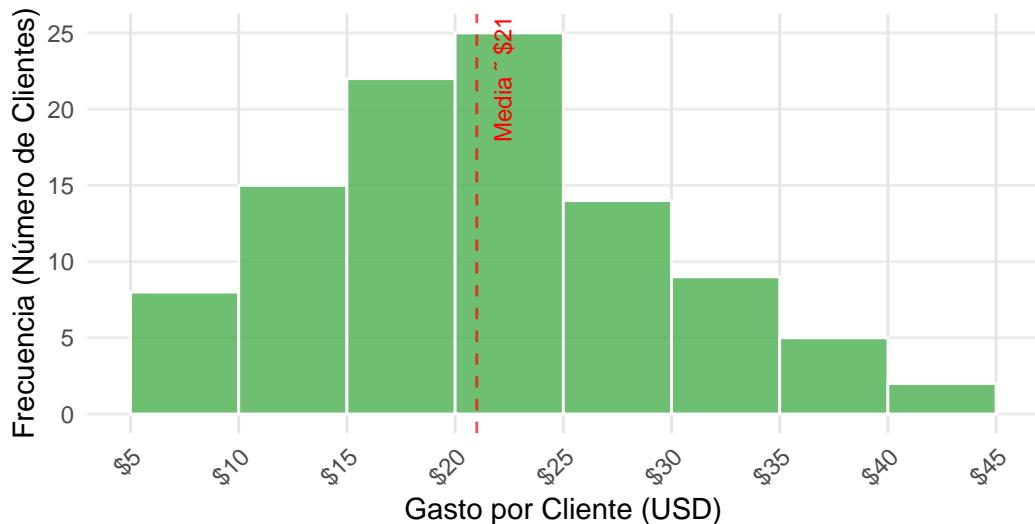
- **Eje X:** Intervalos de valores (clases) de la variable numérica
- **Eje Y:** Frecuencia (cantidad de observaciones en cada intervalo)
- **Barras conectadas:** Sin espacios, indicando continuidad de los datos
- **Forma de distribución:** Permite identificar patrones como simetría, sesgo o multimodalidad

El histograma nos permite visualizar rápidamente:

- **¿Dónde se concentran los datos?** (moda o pico más alto)

Distribución de Gastos por Cliente

Histograma basado en 100 observaciones



Los datos muestran una distribución aproximadamente normal con sesgo ligero hacia la derecha

Figura 2.5: Histograma - Distribución de Gastos por Cliente

- ¿Cómo se distribuyen? (simétrica, sesgada a la izquierda/derecha)
- ¿Hay valores atípicos? (barras aisladas en los extremos)

i Interpretación del histograma

Forma de la distribución:

- **Aproximadamente normal:** La distribución tiene forma de campana con un pico central
- **Sesgo ligero:** Hay una “cola” más larga hacia la derecha (valores altos)
- **Concentración:** La mayoría de clientes gasta entre \$15-\$30

Información práctica:

- **Valor típico:** Alrededor de \$20-\$25 (pico del histograma)
- **Rango común:** El 70 % de clientes gasta menos de \$25
- **Valores extremos:** Muy pocos clientes gastan más de \$35

¿Cuándo usar un histograma vs. gráfico de barras?

Tabla 2.5: Histograma vs. Gráfico de Barras

| Criterio | Histograma | Gráfico de Barras |
|-----------------------------|-------------------------------------|-------------------------------------|
| Tipo de datos | Numéricos continuos | Categóricos |
| Separación de barras | Sin espacios (datos continuos) | Con espacios (categorías distintas) |
| Objetivo principal | Mostrar distribución de frecuencias | Comparar categorías |
| Ejemplo | Edades, pesos, tiempos | Colores, marcas, géneros |

2.2.4.3. Gráfico de Dispersión

El **gráfico de dispersión** (scatter plot) es la herramienta principal para visualizar la **relación entre dos variables numéricas**. A diferencia del histograma que muestra una sola variable, el gráfico de dispersión permite explorar si existe algún patrón, tendencia o correlación entre dos mediciones diferentes.

Sigamos con el restaurante: además del **gasto de cada cliente** (en dólares), registramos el **tiempo que permanecen** en el establecimiento (en minutos). Queremos saber si existe alguna relación entre estas dos variables continuas.

En el gráfico de dispersión: - Cada **punto** representa a un cliente individual. - El **eje X** muestra el tiempo de permanencia (en minutos). - El **eje Y** muestra el gasto total de ese cliente (en dólares).

Patrones que puedes identificar:

💡 Tipos de relaciones en un gráfico de dispersión

- **Correlación positiva:** Los puntos forman una nube que asciende (). A mayor X, mayor Y
- **Correlación negativa:** Los puntos forman una nube que desciende (). A mayor X, menor Y
- **Sin correlación:** Los puntos se distribuyen aleatoriamente sin patrón claro
- **Correlación curvilínea:** Los puntos siguen una curva (ej. forma de U o parábola)

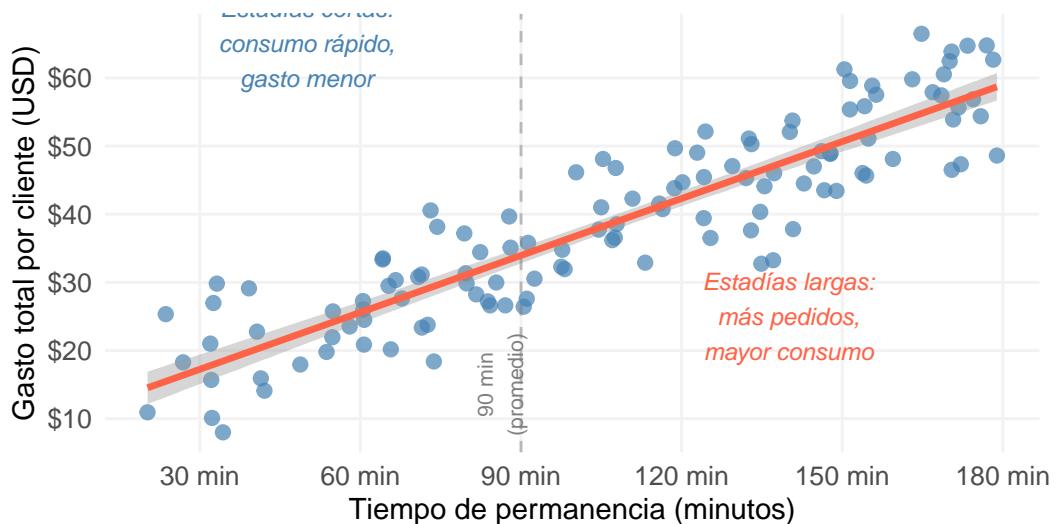
Interpretación del gráfico de dispersión:

El ejemplo muestra una **correlación positiva** entre el tiempo de permanencia y el gasto del cliente. Esta relación tiene sentido desde la perspectiva del negocio:

- **Tendencia ascendente:** Los clientes que permanecen más tiempo tienden a gastar más

Tiempo de Permanencia vs. Gasto del Cliente

Análisis de 120 clientes en el restaurante



Correlación positiva: mayor tiempo de estadía se asocia con mayor gasto

Figura 2.6: Gráfico de Dispersion - Relación entre Tiempo de Permanencia y Gasto del Cliente

- **Relación lógica:** Más tiempo permite más pedidos (aperitivos, postres, bebidas adicionales)
- **Variabilidad natural:** No todos los puntos siguen la línea perfectamente - algunos clientes gastan mucho en poco tiempo (pedidos caros) y otros gastan poco aunque permanezcan mucho tiempo
- **Rango de comportamientos:** Clientes con estadías cortas (20-60 min) gastan típicamente \$8-\$30, mientras que los de estadías largas (120-180 min) gastan frecuentemente \$35-\$70

Insights para el negocio:

- **Estrategia de retención:** Mantener a los clientes más tiempo puede **incrementar las ventas**
- **Ambiente acogedor:** Espacios cómodos que inviten a quedarse más tiempo
- **Menú estratégico:** Ofrecer aperitivos, postres y bebidas para **extender la experiencia**
- **Identificar oportunidades:** Clientes con estadías largas pero bajo gasto podrían necesitar más atención del mesero

¿Cuándo usar un gráfico de dispersión?

Ideal para:

- Explorar relaciones entre dos variables numéricas
- Identificar correlaciones positivas, negativas o ausencia de correlación
- Detectar valores atípicos (outliers)
- Validar supuestos antes de análisis estadísticos más avanzados

Limitaciones:

- Solo muestra dos variables a la vez
- La correlación no implica causalidad
- Puede ser difícil interpretar con muchos puntos sobrepuertos

3 Explorando tus Datos con Estadísticos Descriptivos

Imagina que eres gerente de un restaurante que acaba de lanzar un nuevo menú y quieres evaluar la satisfacción de tus clientes a partir de las calificaciones que te han dado. Tienes cientos de puntuaciones y necesitas entender rápidamente cuál es la opinión general, qué tan variadas son las experiencias y si hay patrones o valores atípicos que merecen atención.

¿Cómo puedes resumir y comprender toda esa información de forma clara y sencilla? Para eso utilizamos los estadísticos descriptivos, herramientas fundamentales que nos permiten condensar grandes cantidades de datos en medidas simples y significativas.

Estas medidas serán esenciales para tomar decisiones informadas que mejoren tu decisión.

3.1. Estadísticos Descriptivos

Los estadísticos descriptivos son herramientas clave que nos permiten resumir y comprender mejor la información contenida en nuestros datos. Para facilitar su estudio, los podemos clasificar en cuatro categorías principales:

- **Medidas de tendencia central:** Estas nos muestran el valor típico o representativo en un conjunto de datos, es decir, alrededor de qué número se agrupan la mayoría de las observaciones.
- **Medidas de variación:** Nos indican qué tan dispersos o concentrados están los datos respecto a esa tendencia central, ayudándonos a entender la consistencia o diversidad dentro de la información.
- **Medidas de forma:** Describen la distribución general de los datos, revelando si están simétricos, sesgados hacia un lado o presentan picos o colas particulares.
- **Medidas de relación:** Evalúan cómo dos variables numéricas se relacionan entre sí, especialmente si existe una conexión lineal que pueda ser útil para análisis más avanzados.

3.2. Medidas de Tendencia Central

Cuando queremos entender las calificaciones que los clientes dan a tu restaurante, buscamos encontrar un valor que represente lo que la mayoría piensa. Las medidas de tendencia central nos ayudan a esto.

3.2.1. Media (Promedio)

La media o promedio es una forma simple pero útil de resumir un conjunto de datos. La media es el valor que obtienes al sumar todas las calificaciones y dividirlas entre el número total de clientes.

💡 Matemáticamente se escribiría así:

$$\text{Media} = \bar{x} = \frac{x_1 + x_2 + \dots + x_n}{n}$$

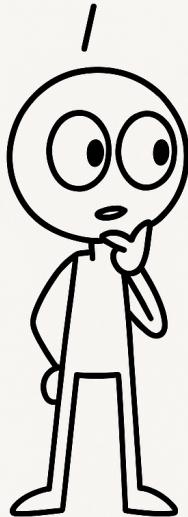
Por ejemplo, si cinco clientes dieron las calificaciones: 4, 5, 3, 4 y 5, la media será:

💡

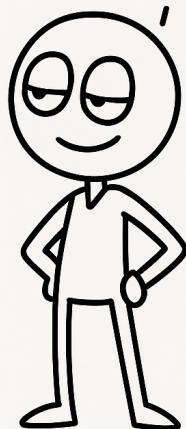
$$\bar{x} = \frac{4 + 5 + 3 + 4 + 5}{5} = \frac{21}{5} = 4.2$$

Sin embargo la media, tiene limitaciones: no muestra cómo están distribuidos los datos y puede ser engañosa si hay grandes diferencias entre los valores. Por ejemplo:

Cuál sería mi salario inicial?



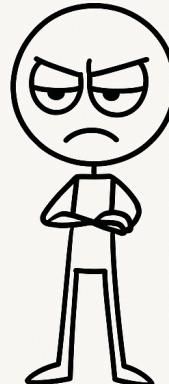
Pongámoslo así:
el promedio de
salario inicial es
20 millones.



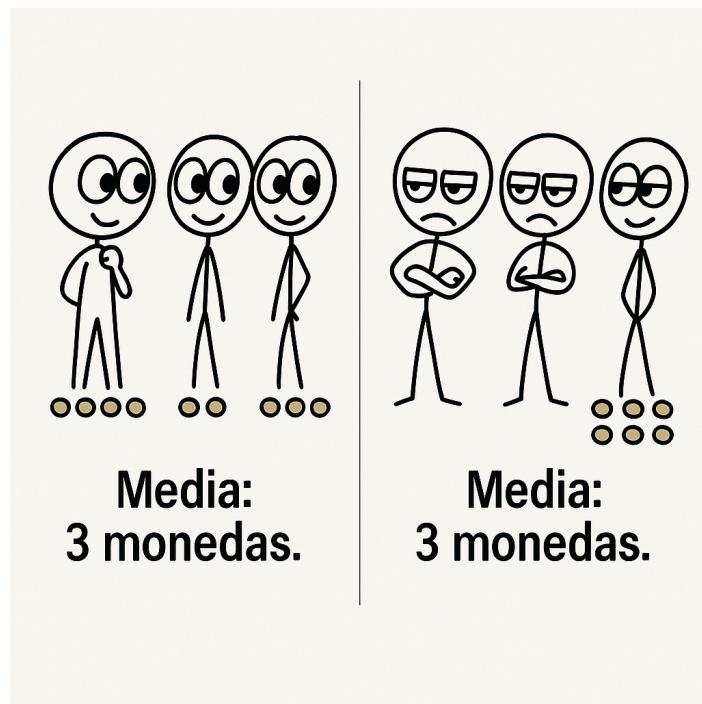
tú → 5 millones

com-pañeros { 5 millones
5 millones
5 millones
couniones
5 millones
5 millones
hijo del CEO 125 millones

Promedio:
20 millones



Aquí hay otro ejemplo donde la media es la misma en ambas situaciones pero el contexto individual que esconde es muy diferente:



3.2.2. Mediana

Volviendo a las calificaciones de tu restaurante, la **mediana** es el valor que está justo en el medio cuando ordenas todas las opiniones de menor a mayor. Esto significa que la mitad de los clientes dieron una calificación igual o menor que la mediana, y la otra mitad dio una calificación igual o mayor.

Por ejemplo, si tus clientes calificaron así: 3, 4, 4, 5, 5, la mediana es 4 — porque es el valor que divide el grupo en dos partes iguales.



Mediana = valor central en datos ordenados

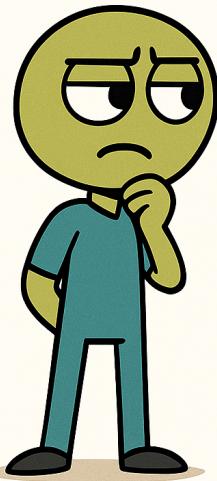
Una gran ventaja de la mediana es que **no se ve afectada por calificaciones muy bajas o muy altas** que podrían distorsionar la media. Por ejemplo, si alguien puso un 1 o un 10, la mediana sigue mostrando el punto medio real de la mayoría.

Sin embargo, la mediana **no nos dice qué tan dispersas están las calificaciones a cada lado**. Por eso, para entender mejor la variabilidad de las opiniones, necesitaremos otras medidas que veremos más adelante.

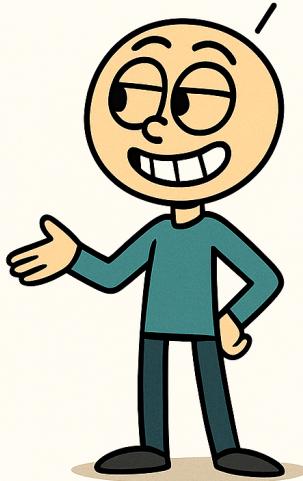
A continuación podemos ver un ejemplo donde la mediana es usada para entregar un mensaje erróneo:

Entonces,
por qué
debería
invertir
contigo?

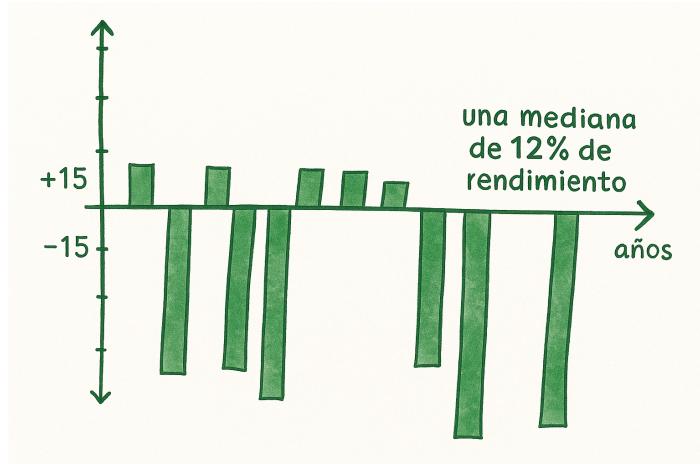
I



No es por
presumir, pero
mi fondo tiene
una **mediana**
de **12%** de
rendimiento al año



Pero, los rendimientos anuales del fondo:



Miremos otro ejemplo donde la mediana similar no implica datos similares. Siempre hay tener una combinación de datos para tomar decisiones correctas



3.2.3. Moda

La **moda** es la calificación que más se repite entre tus clientes. Es como la opinión más común o popular sobre tu restaurante.

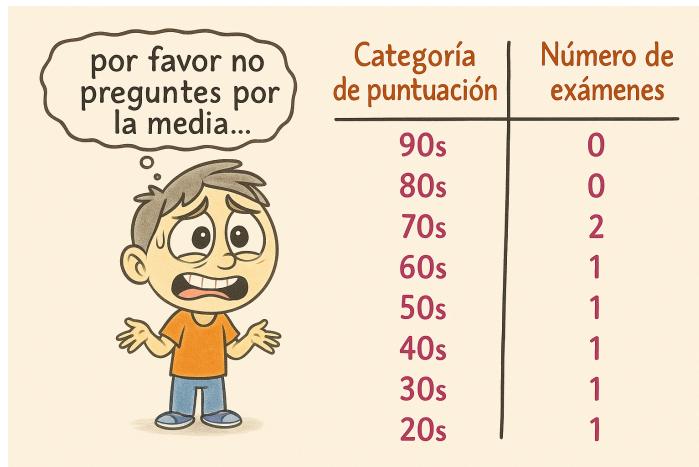
Por ejemplo, si las calificaciones de tus clientes fueron: 3, 4, 4, 5, 5, tanto 4 como 5 se repiten dos veces, por lo que hay dos modas: 4 y 5.

Si no hay repeticiones exactas, se pueden agrupar en categorías y tomar como moda la más común. Es útil especialmente con datos no numéricos, como colores o preferencias políticas, donde no tiene sentido calcular promedios.



Moda = valor que aparece con mayor frecuencia

Su limitación: no considera la totalidad ni la distribución de los datos, y lo más común no siempre es lo más representativo. Miremos este ejemplo:



3.3. Medidas de Variación

Las medidas de variación nos ayudan a entender qué tan diferentes o dispersos están los datos entre sí. Es decir, nos dicen si las opiniones o valores están muy juntos o muy separados.

3.3.1. Rango

Las medidas de variación nos ayudan a entender qué tan diferentes o dispersos están los datos entre sí. Es decir, nos dicen si las opiniones o valores están muy juntos o muy separados.

Por ejemplo, si las calificaciones en tu restaurante van desde 2 hasta 5, el rango sería:



$$\text{Rango} = 5 - 2 = 3$$

Esto nos dice que las opiniones varían en un rango de 3 puntos, desde una calificación baja hasta una alta.

Su principal ventaja es su simplicidad, da una idea rápida del “ancho” del conjunto de datos.

Pero su debilidad es igual de clara, solo considera los valores extremos, ignorando por completo todos los datos intermedios.

Miremos un ejemplo de un rango que da una impresión incorrecta:

