Resumen de Probabilidad y Estadística

Este resumen está diseñado para ayudarte a entender los conceptos clave de probabilidad y estadística de forma clara y práctica. Cada tema incluye definiciones, explicaciones detalladas, ejemplos resueltos y, en algunos casos, gráficos para que visualices mejor las ideas. ¡Empecemos!

1. Representación de datos a través de tablas y gráficos

La estadística es una herramienta poderosa que nos permite recolectar, organizar, analizar y presentar datos para tomar decisiones informadas. Por ejemplo, podemos estudiar las notas de un curso, las alturas de un grupo de personas o las preferencias de los consumidores. La representación de datos en tablas y gráficos facilita su comprensión.

Definiciones preliminares:

- **Población**: Es el conjunto completo de elementos que comparten una característica que queremos estudiar. Por ejemplo, si queremos saber la edad promedio de los estudiantes en un país, la población son *todos* los estudiantes de ese país.
- **Muestra**: Es un subconjunto de la población, seleccionado para representarla. Por ejemplo, si tomamos 100 estudiantes de diferentes escuelas, esa es nuestra muestra. Una buena muestra debe ser representativa para que las conclusiones sean válidas.
- **Variable cualitativa**: Describe características no numéricas. Hay dos tipos:
  - **Ordinal**: Tiene un orden, como el nivel educativo (primaria, secundaria, universidad).
  - **Nominal**: No tiene orden, como colores favoritos o nombres.
- **Variable cuantitativa**: Representa valores numéricos. Puede ser:
  - **Continua**: Puede tomar cualquier valor en un rango, como el peso o la altura.
  - **Discreta**: Solo toma valores enteros, como el número de hijos o los autos en un estacionamiento.

Ejemplo 1 (Definiciones): En un estudio sobre estudiantes de un colegio:

- Población: Todos los estudiantes del colegio.
- Muestra: 50 estudiantes seleccionados al azar.
- Variable cualitativa nominal: Color favorito (rojo, azul, verde).
- Variable cualitativa ordinal: Nivel de satisfacción (bajo, medio, alto).
- Variable cuantitativa continua: Altura en centímetros.
- Variable cuantitativa discreta: Número de hermanos.

Tablas de datos no agrupados: Cuando los datos son pocos o tienen valores específicos, usamos tablas de datos no agrupados. Estas tablas organizan la información de la siguiente manera:

- **Variable**: La característica que medimos (ejemplo: peso).
- **Dato** (x_i): Cada valor observado de la variable.
- **Frecuencia absoluta** (f): Cuántas veces aparece un dato.
- **Frecuencia absoluta acumulada** (F): La suma de las frecuencias hasta un dato.
- **Frecuencia relativa** (fr): La proporción de un dato respecto al total, calculada como $fr = \frac{f}{n}$, donde n es el número total de datos. Se expresa como fracción o porcentaje.
- **Frecuencia relativa acumulada** (Fr): La suma de las frecuencias relativas hasta un dato.
- **Rango**: La diferencia entre el mayor y el menor dato.
- **Tamaño de muestra** (n): El número total de datos.

Ejemplo 2 (Tabla de datos no agrupados): Supongamos que medimos el número de libros leídos en un mes por 10 estudiantes: 2, 3, 2, 4, 3, 2, 5, 3, 4, 2. Construyamos la tabla:

Libros (x_i)	f	F	fr	Fr
2	4	4	$\frac{4}{10} = 40\%$	40%
3	3	7	$\frac{3}{10} = 30\%$	70%
4	2	9	$\frac{2}{10} = 20\%$	90%
5	1	10	$\frac{1}{10} = 10\%$	100%
Total (n)		10	100%	

Explicación:

- f: Contamos cuántas veces aparece cada número de libros.
- F: Sumamos las frecuencias acumuladas (4, 4+3=7, 7+2=9, 9+1=10).
- fr: Dividimos f entre $n = 10$ y lo expresamos como porcentaje.
- Fr: Sumamos las frecuencias relativas acumuladas.
- Rango: $5 - 2 = 3$.

Tablas de datos agrupados: Cuando tenemos muchos datos o valores continuos, agrupamos los datos en intervalos. Estas tablas incluyen:

- **Intervalo**: Rango de valores, con un límite inferior y superior (ejemplo: $[150, 160[$ significa $150 \leq x < 160$).
- **Amplitud**: La diferencia entre el límite superior e inferior del intervalo.
- **Marca de clase** (MC): El promedio de los límites del intervalo, que representa al intervalo.
- **Frecuencia absoluta** (f): Cuántos datos caen en cada intervalo.

Ejemplo 3 (Tabla de datos agrupados): Medimos las alturas (en cm) de 20 estudiantes: 152, 155, 160, 162, 165, 168, 170, 172, 175, 178, 153, 157, 161, 164, 166, 169, 171, 174, 176, 179. Usamos intervalos de amplitud 10 cm: $[150, 160[$, $[160, 170[$, $[170, 180[$. La tabla es:

Intervalo	MC	f
$[150, 160[$	$\frac{150+160}{2} = 155$	5
$[160, 170[$	$\frac{160+170}{2} = 165$	7
$[170, 180[$	$\frac{170+180}{2} = 175$	8
	Total	20

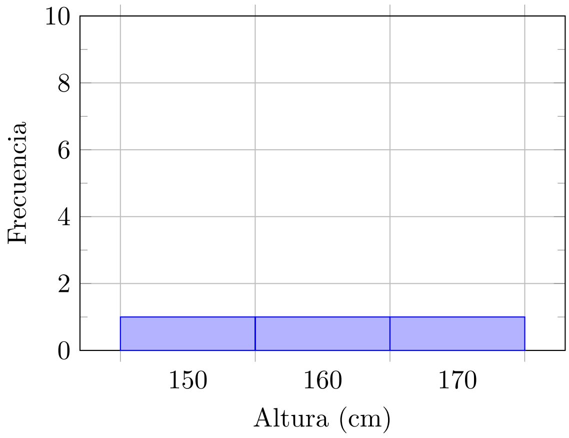
Explicación:

- Intervalos: Dividimos los datos en rangos de 10 cm.
- Marca de clase: Calculamos el promedio de los límites.
- Frecuencia: Contamos cuántos datos caen en cada intervalo (por ejemplo, 152, 153, 155, 157, 161 están en $[150, 160[$).

Representaciones gráficas: Los gráficos nos ayudan a visualizar los datos de forma rápida y clara. Los más comunes son:

- **Histograma**: Usa barras para datos agrupados en intervalos. Las barras están pegadas porque los datos son continuos.
- **Gráfico de barras**: Usa barras separadas para variables discretas o cualitativas.
- **Gráfico de líneas**: Compara datos de varias poblaciones a lo largo del tiempo o categorías.
- **Gráfico circular**: Muestra proporciones como porcentajes de un todo.

Ejemplo 4 (Gráficos): Usemos los datos de la tabla de alturas para crear un histograma.



Explicación: El histograma muestra las frecuencias de las alturas en los intervalos $[150, 160[$, $[160, 170[$, $[170, 180[$. Las barras están unidas porque los datos son continuos.

Ejemplo 5 (PAES - Frecuencias relativas): En una tabla de edades de 40 personas, con intervalos $[10, 20[$, $[20, 30[$, $[30, 40[$, las frecuencias son 12, 18, 10, y las frecuencias relativas son 30%, 45%, 25%. ¿Qué porcentaje de personas tiene menos de 30 años?

- Las frecuencias relativas para $[10, 20[$ y $[20, 30[$ son 30% y 45%.
- Suma: $30\% + 45\% = 75\%$.
- Respuesta: El 75% de las personas tiene menos de 30 años.

2. Medidas de tendencia central y rango

Las medidas de tendencia central resumen un conjunto de datos con un valor representativo. El rango mide la dispersión de los datos.

Para datos no agrupados:

- **Media**: Es el promedio aritmético, calculado como $\bar{x} = \frac{x_1+x_2+\cdots+x_n}{n}$. Representa el "centro" de los datos.
- **Moda**: Es el valor que aparece con mayor frecuencia. Puede haber más de una moda (bimodal, multimodal) o ninguna.
- **Mediana**: Es el valor central cuando los datos están ordenados. Si n es par, se promedian los dos valores centrales.
- **Rango**: Es la diferencia entre el valor máximo y el mínimo.

Ejemplo 6 (Datos no agrupados): Considera las notas de un estudiante: 4, 5, 6, 4, 7. Calcula la media, moda, mediana y rango.

- Media: $\bar{x} = \frac{4+5+6+4+7}{5} = \frac{26}{5} = 5,2$.
- Moda: El 4 aparece dos veces, los demás una vez. Moda = 4.
- Mediana: Ordenamos: 4, 4, 5, 6, 7. La mediana es 5 (tercer valor).
- Rango: $7 - 4 = 3$.
- Respuesta: Media = 5,2; Moda = 4; Mediana = 5; Rango = 3.

Para datos en tabla sin intervalos: Cuando los datos están en una tabla de frecuencias, usamos las frecuencias para calcular las medidas:

- **Media**: $\bar{x} = \frac{x_1 f_1+x_2 f_2+\cdots+x_k f_k}{n}$, donde $n = f_1 + f_2 + \cdots + f_k$.
- **Moda**: El dato con la mayor frecuencia absoluta (f).
- **Mediana**: El dato cuya frecuencia acumulada (F) alcanza o supera la posición $\frac{n}{2}$.

Ejemplo 7 (Tabla sin intervalos): Usemos la tabla del Ejemplo 2 (número de libros leídos):

Libros (x_i)	f	F
2	4	4
3	3	7
4	2	9
5	1	10

Calcula la media, moda y mediana.

- Media: $\bar{x} = \frac{(2\cdot4)+(3\cdot3)+(4\cdot2)+(5\cdot1)}{10} = \frac{8+9+8+5}{10} = \frac{30}{10} = 3.$
- Moda: El valor con mayor f es 2 ($f = 4$). Moda = 2.
- Mediana: $n = 10$, la posición central es $\frac{10}{2} = 5$. Según F, el quinto dato está en $x_i = 3$ ($F = 7$ supera 5). Mediana = 3.
- Respuesta: Media = 3; Moda = 2; Mediana = 3.

Para datos con intervalos: Para datos agrupados en intervalos, usamos las marcas de clase:

- **Media**: $\bar{x} = \frac{\mathrm{MC}_1 f_1 + \mathrm{MC}_2 f_2 + \cdots + \mathrm{MC}_k f_k}{n}$.
- **Moda**: El intervalo con la mayor frecuencia absoluta.
- **Mediana**: El intervalo que contiene la posición $\frac{n}{2}$. Para mayor precisión, se puede usar interpolación lineal:

$$\text{Mediana} = L_i + \left(\frac{\frac{n}{2} - F_{i-1}}{f_i}\right) \cdot h$$

  donde L_i es el límite inferior del intervalo mediano, F_{i-1} es la frecuencia acumulada anterior, f_i es la frecuencia del intervalo, y h es la amplitud.

Ejemplo 8 (Datos con intervalos): Usemos la tabla del Ejemplo 3 (alturas):

Intervalo	MC	f
$[150, 160[$	155	5
$[160, 170[$	165	7
$[170, 180[$	175	8

Calcula la media, moda y mediana.

- Media: $\bar{x} = \frac{(155\cdot5)+(165\cdot7)+(175\cdot8)}{20} = \frac{775+1155+1400}{20} = \frac{3330}{20} = 166,5$,cm.
- Moda: El intervalo con mayor f es $[170, 180[$ ($f = 8$). Moda = $[170, 180[$.
- Mediana: $n = 20$, $\frac{n}{2} = 10$. El décimo dato está en el intervalo $[160, 170[$, ya que $F = 5 + 7 = 12$ supera 10. Usamos interpolación:

$$\text{Mediana} = 160 + \left(\frac{10-5}{7}\right) \cdot 10 = 160 + \frac{5}{7} \cdot 10 \approx 160 + 7,14 \approx 167,14\text{,cm.}$$

- Respuesta: Media = 166,5 cm; Moda = $[170, 180[$; Mediana $\approx$ 167,14 cm.

Ejemplo 9 (PAES - Comparación de medias y medianas): Notas de dos estudiantes, Ana y Luis:

Asignatura	Ana	Luis
Matemáticas	6,0	5,5
Lenguaje	5,5	6,0
Ciencias	4,8	4,5
Historia	6,2	5,0
Inglés	5,0	5,8

Calcula las medias (P_A, P_L) y medianas (R_A, R_L).

- Media de Ana: $P_A = \frac{6,0+5,5+4,8+6,2+5,0}{5} = \frac{27,5}{5} = 5,5$.
- Media de Luis: $P_L = \frac{5,5+6,0+4,5+5,0+5,8}{5} = \frac{26,8}{5} = 5,36$.
- Mediana de Ana: Ordenamos: 4,8, 5,0, 5,5, 6,0, 6,2. Mediana $R_A = 5,5$.
- Mediana de Luis: Ordenamos: 4,5, 5,0, 5,5, 5,8, 6,0. Mediana $R_L = 5,5$.
- Comparación: $P_A > P_L$ (5,5 ¿ 5,36), $R_A = R_L$ (5,5 = 5,5).

3. Medidas de posición

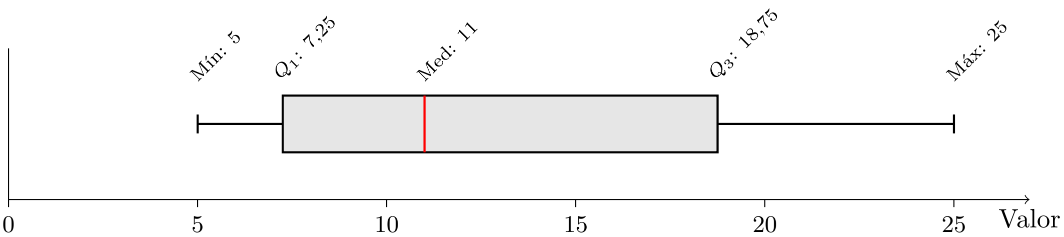
Las medidas de posición nos ayudan a entender cómo se distribuyen los datos dividiendo el conjunto en partes. Son útiles para comparar datos o identificar valores extremos.

- **Percentil** (P_X): El valor por debajo del cual está el X% de los datos. Por ejemplo, P_{90} significa que el 90% de los datos son menores o iguales a ese valor.
- **Cuartiles**: Dividen los datos en cuatro partes iguales:
  - $Q_1 = P_{25}$: 25% de los datos son menores.
  - $Q_2 = P_{50}$: Es la mediana.
  - $Q_3 = P_{75}$: 75% de los datos son menores.
- **Boxplot** (diagrama de caja): Representa gráficamente el mínimo, Q_1, Q_2, Q_3 y el máximo. Es útil para visualizar la dispersión y detectar valores atípicos.

Ejemplo 10 (Percentiles y cuartiles): Considera los datos: 5, 7, 8, 10, 12, 15, 20, 25. Calcula Q_1, Q_2, Q_3.

- Ordenados: 5, 7, 8, 10, 12, 15, 20, 25.
- $n = 8$, posiciones:
  - Q_1 (posición $0,25 \cdot (8+1) = 2,25$): Interpolamos entre el segundo y tercer dato: $7 + 0,25 \cdot (8-7) = 7,25$.
  - Q_2 (posición $0,5 \cdot (8+1) = 4,5$): Promedio de cuarto y quinto dato: $\frac{10+12}{2} = 11$.
  - Q_3 (posición $0,75 \cdot (8+1) = 6,75$): Interpolamos entre el sexto y séptimo dato: $15 + 0,75 \cdot (20-15) = 15 + 3,75 = 18,75$.
- Respuesta: $Q_1 = 7,25$, $Q_2 = 11$, $Q_3 = 18,75$.

Gráfico (Boxplot Horizontal):



Explicación: El boxplot muestra el mínimo (5), Q_1 (7,25), la mediana (11), Q_3 (18,75) y el máximo (25). La caja representa el rango intercuartílico ($Q_3 - Q_1$).

Ejemplo 11 (PAES - Cuartiles): Conjunto de datos: 1, 3, 4, 6, 8, 10, 12, 15. Hallar Q_3.

- Ordenados: 1, 3, 4, 6, 8, 10, 12, 15.
- $n = 8$, posición de Q_3: $0,75 \cdot (8+1) = 6,75$.
- Interpolamos entre el sexto y séptimo dato: $10 + 0,75 \cdot (12-10) = 10 + 1,5 = 11,5$.
- Respuesta: $Q_3 = 11,5$.

4. Reglas de las probabilidades

La probabilidad mide la posibilidad de que ocurra un evento en un experimento aleatorio. Es un número entre 0 (imposible) y 1 (seguro).

Definiciones preliminares:

- **Experimento aleatorio**: Un proceso con resultados impredecibles, como lanzar un dado.
- **Espacio muestral** (E): El conjunto de todos los resultados posibles. Por ejemplo, al lanzar un dado, $E = \{1, 2, 3, 4, 5, 6\}$.
- **Cardinalidad**: El número de elementos en E (en el dado, $|E| = 6$).
- **Evento**: Un subconjunto de E. Por ejemplo, sacar un número par: $\{2, 4, 6\}$.
- **Evento cierto**: Ocurre siempre (probabilidad = 1). Ejemplo: Sacar un número menor a 7 en un dado.
- **Evento imposible**: Nunca ocurre (probabilidad = 0). Ejemplo: Sacar un 7 en un dado.
- **Eventos mutuamente excluyentes**: No pueden ocurrir al mismo tiempo. Ejemplo: Sacar un 1 y un 2 en un solo lanzamiento.
- **Eventos complementarios**: Son opuestos y su unión cubre todo E. Ejemplo: Sacar par y sacar impar.

Ejemplo 12 (Definiciones de probabilidad): Al lanzar una moneda, describe el espacio muestral y algunos eventos.

- Espacio muestral: $E = \{\text{cara}, \text{cruz}\}$.
- Cardinalidad: $|E| = 2$.
- Evento: Sacar cara = $\{\text{cara}\}$.
- Evento cierto: Sacar cara o cruz = E.
- Evento imposible: Sacar un número = $\emptyset$.
- Eventos mutuamente excluyentes: Sacar cara y sacar cruz.
- Eventos complementarios: Sacar cara y sacar cruz.

Probabilidad clásica (Laplace): Si todos los resultados son igualmente probables, la probabilidad de un evento A es:

$$P(A) = \frac{\text{número de casos favorables}}{\text{número de casos totales}}.$$

Ejemplo 13 (Probabilidad clásica): Al lanzar un dado, ¿cuál es la probabilidad de sacar un número mayor a 4?

- Espacio muestral: $E = \{1, 2, 3, 4, 5, 6\}$, $|E| = 6$.
- Evento A: Sacar un número mayor a 4, $A = \{5, 6\}$.
- $P(A) = \frac{\text{casos favorables}}{\text{casos totales}} = \frac{2}{6} = \frac{1}{3}$.
- Respuesta: $\frac{1}{3}$.

Ley de los grandes números: Si repetimos un experimento muchas veces, la proporción de veces que ocurre un evento se aproxima a su probabilidad teórica. Por ejemplo, al lanzar una moneda muchas veces, la proporción de caras se acercará a $\frac{1}{2}$.

Ejemplo 14 (Ley de los grandes números): Lanzas una moneda 10 veces y obtienes 7 caras. ¿Es esto consistente con $P(\text{cara}) = \frac{1}{2}$?

- Con solo 10 lanzamientos, 7 caras (70%) no es inusual, ya que las muestras pequeñas pueden variar.
- Si lanzas la moneda 1000 veces y obtienes 700 caras, esto sería menos probable, ya que la proporción debería acercarse a 50%.

- Explicación: La ley de los grandes números asegura que con más repeticiones, los resultados se estabilizan cerca de la probabilidad teórica.

Suma de probabilidades:

- **Eventos mutuamente excluyentes**: $P(A \cup B) = P(A) + P(B)$. Ejemplo: La probabilidad de sacar un 1 o un 2 en un dado.
- **Eventos no mutuamente excluyentes**: $P(A \cup B) = P(A) + P(B) - P(A \cap B)$. Ejemplo: Sacar una carta que sea rey o de corazones.

Ejemplo 15 (Suma de probabilidades): En un mazo de 52 cartas, calcula la probabilidad de sacar una carta que sea rey o de corazones.

- $P(\text{rey}) = \frac{4}{52}$, ya que hay 4 reyes.
- $P(\text{corazones}) = \frac{13}{52}$, ya que hay 13 cartas de corazones.
- $P(\text{rey} \cap \text{corazones}) = \frac{1}{52}$, ya que hay 1 rey de corazones.
- $P(\text{rey} \cup \text{corazones}) = \frac{4}{52} + \frac{13}{52} - \frac{1}{52} = \frac{16}{52} = \frac{4}{13}$.
- Respuesta: $\frac{4}{13}$.

Producto de probabilidades: Para eventos independientes (donde uno no afecta al otro), $P(A \cap B) = P(A) \cdot P(B)$. Ejemplo: Sacar cara en dos lanzamientos de moneda.
Ejemplo 16 (Producto de probabilidades): Lanzas dos dados. ¿Cuál es la probabilidad de sacar un 6 en ambos?

- $P(6 \text{ en el primer dado}) = \frac{1}{6}$.
- $P(6 \text{ en el segundo dado}) = \frac{1}{6}$.
- Como los lanzamientos son independientes: $P(6 \text{ en ambos}) = \frac{1}{6} \cdot \frac{1}{6} = \frac{1}{36}$.
- Respuesta: $\frac{1}{36}$.

Ejemplo 17 (PAES - Probabilidad): En una urna hay 50 bolas: azules y blancas. La probabilidad de sacar una bola blanca es $\frac{3}{5}$. ¿Cuántas bolas azules hay?

- $P(\text{blanca}) = \frac{3}{5}$, luego $P(\text{azul}) = 1 - \frac{3}{5} = \frac{2}{5}$.
- Total de bolas $= 50$. Bolas azules: $\frac{2}{5} \cdot 50 = 20$.
- Respuesta: 20 bolas azules.