

## Estadística Descriptiva

### 4.2. Organización de Datos en el Ámbito Legal

Después de recopilar nuestros datos (los elementos, variables y sus escalas de medición que ya dominan), el siguiente paso crucial es organizarlos y resumirlos. Esto nos permite ver patrones, tendencias y características clave de un vistazo.

#### 4.2.1. Tablas de Frecuencia

Una **tabla de frecuencia** es una herramienta simple pero poderosa para organizar datos cualitativos o cuantitativos (cuando son pocos valores distintos). Nos muestra cuántas veces aparece cada categoría o valor en nuestro conjunto de datos.

- **Ejemplo en el Derecho:**
  - **Tipos de Delitos en un Juzgado Penal**

Imaginemos que, en un Juzgado Penal de Riobamba, recopilamos los tipos de delitos más comunes registrados en el último mes.

Tipo de Delito	Conteo (Frecuencia Absoluta)	Frecuencia Relativa (%)
Hurto	45	30%
Lesiones	30	20%
Estafa	25	16.7%
Robo	20	13.3%

Tipo de Delito	Conteo (Frecuencia Absoluta)	Frecuencia Relativa (%)
Otros	30	20%
<b>Total</b>	<b>150</b>	<b>100%</b>

- **¿Qué nos dice esto?** Rápidamente vemos que el hurto es el delito más frecuente (30%) en ese juzgado durante el mes, seguido de lesiones y "otros" (que podrían requerir un análisis más detallado de esas categorías). ¡Esto es vital para la asignación de recursos o la prevención del delito!

#### 4.2.2. Tablas de Distribución de Frecuencias (para Datos Agrupados)

Cuando tenemos una gran cantidad de datos cuantitativos con muchos valores distintos (por ejemplo, edades, montos de multas), agruparlos en **intervalos o clases** nos facilita la visualización. Aquí usamos una **tabla de distribución de frecuencias**.

- **Ejemplo en el Derecho: Edades de Infractores de Tránsito**

Supongamos que la Agencia de Tránsito de Riobamba analiza las edades de 200 conductores que cometieron infracciones graves en el último trimestre. En lugar de listar cada edad individual, las agrupamos:

Rango de Edad (años)	Frecuencia Absoluta	Frecuencia Relativa (%)	Frecuencia Acumulada
18 - 25	60	30%	60
26 - 35	80	40%	140

Rango de Edad (años)	Frecuencia Absoluta	Frecuencia Relativa (%)	Frecuencia Acumulada
36 - 45	35	17.5%	175
46 - 55	20	10%	195
56 o más	5	2.5%	200
<b>Total</b>	<b>200</b>	<b>100%</b>	

- **¿Qué nos dice esto?** Observamos que el grupo de edad entre **26 y 35 años** es el que más infracciones graves comete (40%). Además, sabemos que el 70% (30% + 40%) de los infractores tienen 35 años o menos. Esta información puede orientar campañas de educación vial dirigidas a segmentos específicos de la población.

#### 4.2.3. Modelos de Tablas Estadísticas

Los ejemplos anteriores ya son modelos de tablas estadísticas. La clave es que sean **claras, concisas y fáciles de interpretar**. Un buen modelo incluye:

- **Título:** Que explique claramente el contenido.
- **Encabezados de columna:** Nombres claros para cada variable o medida.
- **Unidades de medida:** Si aplica (ej. %, años, \$).
- **Fuente de los datos:** De dónde provienen.
- **Ejemplo en el Derecho: Modelos y su utilidad**
  - **Tabla de Frecuencia Simple:** Ideal para variables cualitativas o cuantitativas discretas con pocas categorías (ej., tipo de delito, género del demandante).

- **Tabla de Frecuencia con Clases (Intervalos):** Perfecta para variables cuantitativas continuas o discretas con muchos valores (ej., duración de juicios, montos de sentencias).

La elección del modelo de tabla depende del tipo de variable y de la cantidad de datos, siempre buscando la máxima claridad para el usuario final (que podría ser un juez, un colega abogado o una autoridad legislativa).

## Caso de ejemplo

### Análisis de Causas de Retraso en la Sala de Audiencias

Imaginemos que la **Presidencia de la Corte Provincial de Justicia de Chimborazo**, aquí en Riobamba, está preocupada por los frecuentes retrasos en el inicio de las audiencias. Esto afecta la eficiencia del sistema judicial, genera inconformidad en los usuarios y abogados, y puede incluso influir en la duración total de los procesos.

Para abordar este problema, han decidido realizar un estudio durante un mes (digamos, agosto de 2025) en una sala de audiencias específica.

#### **Los datos recopilados son los siguientes:**

El equipo observó **120 audiencias programadas** durante el mes. Para cada audiencia, registraron dos tipos de información:

1. **Causa Principal del Retraso:** Anotaron la razón principal si la audiencia se retrasó, o "Ninguno" si inició a tiempo. Las categorías observadas fueron:
  - Ausencia del Abogado
  - Ausencia de Testigo/Perito
  - Problemas Técnicos (sistema, audio, etc.)
  - Demora del Juez
  - Atraso del Detenido/Custodia

- Falta de Documentos
  - Ninguno (inició a tiempo)
2. **Tiempo de Retraso (en minutos):** Si hubo un retraso, registraron la cantidad exacta de minutos desde la hora programada hasta la hora de inicio real. (Si no hubo retraso, el tiempo fue 0). Los tiempos de retraso observados variaron desde 0 hasta 120 minutos.

Ahora, veamos cómo usaríamos las tablas para organizar estos datos:

Aplicación: Tablas de Frecuencia y Distribución de Frecuencias

*1. Tabla de Frecuencia para la Causa Principal del Retraso (Variable Cualitativa)*

Aquí utilizaremos una **tabla de frecuencia simple** para la variable "**Causa Principal del Retraso**". Esta variable es **cualitativa (nominal)**, ya que las causas son categorías sin un orden inherente.

Causa Principal del Retraso	Frecuencia Absoluta (Conteo)	Frecuencia Relativa (%)
Ausencia del Abogado	30	25.0%
Ausencia de Testigo/Perito	15	12.5%
Problemas Técnicos	10	8.3%
Demora del Juez	5	4.2%

Causa Principal del Retraso	Frecuencia Absoluta (Conteo)	Frecuencia Relativa (%)
Atraso del Detenido/Custodia	8	6.7%
Falta de Documentos	2	1.7%
Ninguno (inició a tiempo)	50	41.7%
<b>Total de Audiencias</b>	<b>120</b>	<b>100.0%</b>

### ¿Qué nos dice esta tabla para la Corte de Chimborazo?

- ¡Rápidamente identificamos que el **41.7% de las audiencias inician a tiempo!** Esto es un buen punto de partida.
- Sin embargo, la principal causa de retraso es la "**Ausencia del Abogado**", representando un significativo 25% de los casos. Esto sugiere una necesidad urgente de comunicación, recordatorios o, incluso, posibles sanciones.
- La "**Ausencia de Testigo/Perito**" también es una causa importante (12.5%), lo que podría llevar a la Corte a revisar los procesos de citación y aseguramiento de comparecencias.

### *2. Tabla de Distribución de Frecuencias para el Tiempo de Retraso (Variable Cuantitativa)*

Ahora, para la variable "**Tiempo de Retraso (en minutos)**", que es **cuantitativa (de razón)** y tiene muchos valores posibles, una tabla de distribución de frecuencias agrupada en intervalos es lo más adecuado. Supongamos que los datos se distribuyen así:

Rango de Tiempo de Retraso (minutos)	Frecuencia Absoluta	Frecuencia Relativa (%)	Frecuencia Acumulada
0 (sin retraso)	50	41.7%	50
1 - 15	35	29.2%	85
16 - 30	20	16.7%	105
31 - 60	10	8.3%	115
61 - 90	4	3.3%	119
91 - 120	1	0.8%	120
<b>Total de Audiencias</b>	<b>120</b>	<b>100.0%</b>	

### ¿Qué nos dice esta tabla para la Presidencia de la Corte?

- Reafirmamos que el **41.7% de las audiencias no tienen retraso**. ¡Eso es genial!
- Cuando hay retraso, la mayoría de ellos son **cortos**: el 29.2% de las audiencias se retrasan entre 1 y 15 minutos, y el 16.7% entre 16 y 30 minutos. Esto significa que casi el **46% de todas las audiencias (29.2% + 16.7%) tienen un retraso de 30 minutos o menos**.
- Los retrasos **superiores a una hora (61 minutos o más)** son raros, ocurriendo en menos del 5% de los casos (3.3% + 0.8%). Sin embargo, un solo retraso de hasta 120 minutos (2 horas) es significativo y podría requerir investigación individual.
- La **frecuencia acumulada** es muy útil: podemos ver que el **70.9% de las audiencias (41.7% + 29.2%) se retrasan 15 minutos o menos (o no se retrasan en absoluto)**.

Como pueden observar, con solo dos tablas, hemos transformado un montón de datos brutos en información valiosa y fácil de entender. La Presidencia de la Corte ahora tiene una base sólida para:

- Implementar programas de capacitación para abogados o normativas más estrictas sobre su puntualidad.
- Revisar los protocolos para asegurar la presencia de testigos y peritos.
- Evaluar si los problemas técnicos requieren una mayor inversión en infraestructura.

#### Medidas de Tendencia Central y Dispersión: "El Corazón" y "La Variación" de los Datos

Una vez que tenemos nuestros datos organizados, queremos resumirlos con números que nos digan dónde está el "centro" de nuestros datos y qué tan "dispersos" están.

##### *4.2.4. Medidas de Tendencia Central*

Estas medidas nos dan una idea del valor "típico" o "central" de un conjunto de datos.

- **a) Media (Promedio)**
  - Se calcula sumando todos los valores y dividiendo por el número total de valores. Es la medida más común.
  - **Ejemplo Legal:** Si un abogado lleva 10 casos en un año y la **duración de sus juicios** (en días) fue: 30, 45, 60, 30, 90, 120, 75, 45, 60, 30.
    - $Media = (30 + 45 + 60 + 30 + 90 + 120 + 75 + 45 + 60 + 30)/10 = 585/10 = 58.5$  días.

- **Utilidad:** El promedio de 58.5 días nos da una expectativa de cuánto tiempo duran los juicios de este abogado. Esto puede ser útil para la planificación o para comparar su eficiencia.

- **b) Mediana**

- Es el valor central en un conjunto de datos ordenado de menor a mayor. Deja la mitad de los datos por debajo y la mitad por encima. Es útil cuando hay valores extremos (outliers).
- **Ejemplo Legal:** Si analizamos los **montos de multas por infracciones graves** (en dólares): 100,150, 200,200, 250,300, 350,400, \$1500.
  - Primero ordenamos: 100,150, 200,200, **\*\*250\*\***, 300, 350,400, \$1500.
  - La Mediana es **\$250**.
- **Utilidad:** A pesar de una multa muy alta de 1500, *la mediana de 250* nos dice que la mitad de las multas son de \$250 o menos, lo que podría ser más representativo del "monto típico" si esa multa de \$1500 es una excepción.

- **c) Moda**

- Es el valor que aparece con mayor frecuencia en un conjunto de datos. Es útil para datos cualitativos.
- **Ejemplo Legal:** Si registramos el **Tipo de Delito más común** en un mes (de nuestra tabla anterior): Hurto, Lesiones, Estafa, Robo, Hurto, Lesiones, Hurto...
  - La **Moda** sería **Hurto** (porque aparece más veces).
- **Utilidad:** Nos indica cuál es la categoría más frecuente. Un juez podría querer saber cuál es el tipo de delito que más ve en su sala.

#### 4.2.5. Medidas de Dispersión

Estas medidas nos dicen qué tan "dispersos" o "variados" están los datos alrededor de la medida central. ¿Están todos los valores muy juntos, o hay mucha diferencia entre ellos?

- **a) Rango**

- Es la diferencia entre el valor máximo y el valor mínimo en un conjunto de datos. Es una medida muy simple de la dispersión total.
- **Ejemplo Legal:** Usando los **montos de multas**: Mínimo = 100, *Máximo* = 1500.
  - $\text{Rango} = 1500 - 100 = \$1400$ .
- **Utilidad:** El rango de \$1400 nos indica una gran variabilidad en los montos de las multas, desde muy bajas hasta muy altas.

- **b) Varianza y Desviación Estándar**

- Son las medidas de dispersión más robustas. Nos dicen cuánto se desvían, en promedio, los valores individuales de la media.
  - La **Varianza** es el promedio de las diferencias al cuadrado de cada dato con respecto a la media.
  - La **Desviación Estándar** es la raíz cuadrada de la varianza. Es más fácil de interpretar porque está en las mismas unidades que los datos originales.
- **Ejemplo Legal:** Si la **duración promedio de los juicios** de un abogado es 58.5 días, y su **desviación estándar** es, digamos, 25 días.
- **Utilidad:** Una desviación estándar alta (como 25 días) nos indica que las duraciones de sus juicios varían mucho alrededor de los 58.5 días. Podría tener juicios de 30 días y otros de 120 días. Si la desviación estándar fuera solo de 5 días, significaría que la mayoría de sus juicios duran entre 53.5 y 63.5 días (es decir, son muy consistentes).

- Para un juez o un legislador, una alta desviación estándar en las sentencias de delitos similares podría indicar una falta de uniformidad o predictibilidad en las decisiones judiciales, lo que podría ser motivo de análisis.