ESTADÍSTICA Y PROBABILIDADES

Por Luis Alberto Freire Sánchez

ISTSGA
SAN GABRIEL
INSTITUTO SUPERIOR TECNOLÓGICO

EDICA XXI
Editorial Académica Internacional

Editorial Académica Internacional XXI EDICAXXI S.A.S.

74 pág.: 17x24 cm

Título: Estadísticas y Probabilidades

Luis Alberto Freire Sánchez

Primera Edición 2024

ISBN: 978-9942-8885-7-0

ESTADÍSTICA Y PROBABILIDADES

Por: Luis Alberto Freire Sánchez

Semblanza del Autor

LUIS ALBERTO FREIRE SÁNCHEZ

Ingeniero en Electrónica y Computación, Máster en Ingeniería de Software y Sistemas Informáticos, director de Investigación Desarrollo e Innovación del Instituto Superior Tecnológico San Gabriel.

Tabla de contenidos

Contenido

Dedicatoria

Dedico este trabajo a:

Dios, por darnos fortaleza y ser luz que guía nuestro camino.

Mi querida esposa Magy y a nuestra hija Hannita, ahora que está iniciando sus pasos Dios la bendiga con amor y sabiduría para que nunca deje de aprender.

Mi padre José, por ser ejemplo de esfuerzo y sabiduría.

Mi querida madre Rosa (+) y mi amada hija Emily (+), con la certeza que estarán disfrutando del infinito amor de nuestro Padre celestial.

Agradecimiento

Al Instituto Superior Tecnológico San Gabriel por haber otorgado las facilidades para el desarrollo de estas páginas.

La esencia del progreso de un hombre no se mide únicamente por sus logros, sino por el viaje interior que conlleva superar los desafíos del camino. Cada paso dado nos llena de sabiduría y nos prepara para alcanzar nuevas metas incluso cuando vientos de adversidad soplen en nuestra contra.

Luis Freire Sánchez.

UNIDAD UNO
INTRODUCCIÓN A LA ESTADÍSTICA

1. INTRODUCCIÓN

La estadística es una herramienta valiosa para tomar decisiones, entender noticias e informes que involucren datos. Esta es una rama de las matemáticas encargada de recopilar, analizar e interpretar datos provenientes de experimentos o de la simple observación de fenómenos naturales. Según Barreto-Villanueva (2012), los primeros indicios sobre el uso de la estadística datan de hace 7.000 años con la invención del dado, no obstante, los primeros registros de su estudio como ciencia aparecen en el siglo XVI cuando el polímota Gerolamo Cardano publica su libro sobre juegos de azar *"Liber de ludo aleae"* donde usa el cálculo de probabilidades para analizar los juegos desde una perspectiva sistemática.

Un siglo atrás, el escritor británico Herbert George Wells dijo *"Algún día el conocimiento estadístico será tan necesario para los ciudadanos, como lo es saber leer y escribir*", y tuvo razón puesto que hoy en día los datos estadísticos se encuentran en la mayor parte de las actividades que nos rodea, por ejemplo, cuando miramos el pronóstico del clima en nuestro teléfono, cuando vemos la tabla de posiciones de nuestro equipo deportivo favorito en la televisión, cuando escuchamos las encuestas de intención de votos de los candidatos políticos en la radio, etc., y no se diga en el ámbito profesional, este está presente en todas las ciencias puesto que el avance científico permite realizar estudios en toda área del conocimiento humano.

En este sentido, el reconocido profesor universitario William Edwards Deming insistía que la educación estadística debía de iniciar antes de la preparatoria, y como ejemplo tenía gusto por contar la historia de un niño de 11 años que a su corta edad había desarrollado una tabla de control para monitorear la llegada puntual de su autobús escolar, comentando "*Ha empezado bien su vida*" (Mason, lind y Marchal, 2003).

Para cumplir nuestros objetivos, en el presente texto mostraremos algunas definiciones necesarias para iniciar el curso.

1.1 IMPORTANCIA

Las herramientas estadísticas son fundamentales en una gran variedad de campos, desde la agricultura, la astronomía, la biología, los negocios, las comunicaciones, la economía, la educación, la electrónica, la geología, las ciencias de la salud y muchos otros campos de la ciencia y la ingeniería (Ramachandran, Tsokos, 2009), gracias a ella podemos tener un mejor entendimiento del mundo que nos rodea para tomar decisiones más acertadas. En términos generales se puede describir algunas aplicaciones de la estadística en diferentes áreas:

En las empresas. El análisis estadístico es ampliamente usado por cadenas de tiendas para determinar el mejor momento para lanzar promociones y descuentos, mientras que en la manufactura es muy común su uso para identificar cuellos de botella en la línea de producción a fin de mejorar la eficiencia productiva.

En la medicina. La estadística se utiliza en ensayos clínicos para determinar la eficacia de los medicamentos evaluando sus efectos secundarios y la seguridad de estos, de igual manera es de suma importancia en investigaciones epidemiologia para identificar los factores que influyen en la salud y las enfermedades que pudieran afectar a una población determinada.

En la informática. La información generada por las computadoras puede ser extremadamente extensa, compleja y difícil de interpretar sin las herramientas adecuadas. Un ejemplo común son los datos recogidos por la técnica de minería de datos usado por las organizaciones para describir patrones de consumo en cada uno de los millones de clientes que acceden a sus aplicaciones web como amazon.com, alibaba.com, etc.

En la educación. La estadística permite medir el rendimiento académico de los estudiantes, evaluar la eficacia de los programas educativos y diseñar estrategias de enseñanza más efectivas, además, es esencial para la investigación educativa ya que permite analizar los datos recopilados en estudios y encuestas para obtener conclusiones significativas. En resumen, la estadística es una herramienta clave para mejorar la calidad de la educación y tomar decisiones informadas en el ámbito educativo.

En las ciencias de la tierra. La estadística es una herramienta fundamental en las ciencias de la tierra, ya que permite analizar y comprender datos complejos relacionados con la geología, meteorología, oceanografía y otras disciplinas. Por ejemplo, es bien sabido que las predicciones climatológicas requieren procesar una gran cantidad de muestras estadísticas para obtener una predicción confiable, lo cual es de suma importancia para supervivencia humana, debido a que, si no previniera la llegada de eventos catastróficos como huracanes, sequías, incendios, inundaciones, entre otros, las personas tendrían serios problemas para sobrevivir.

En conclusión, la estadística es una herramienta de suma importancia para comprender el mundo que nos rodea, sin embargo, su uso indebido puede tener consecuencias graves, por tanto, en necesario que los estudiantes de estadística se adhieran a estándares éticos para garantizar que las decisiones basadas en datos sean justas, precisas y confiables. Sin una ética sólida, los datos pueden ser manipulados o malinterpretados, lo que puede llevar a decisiones equivocadas que tienen consecuencias negativas para las personas y las organizaciones involucradas.

1.2 DEFINICIÓN DE ESTADÍSTICA Y SUS TIPOS

Estadística. Se refiere al conjunto de métodos, normas, reglas y principios usados en la planeación de estudios o experimentos para obtener datos, organizarlos, agruparlos, analizarlos, interpretarlos y hallar conclusiones. (Pérez, 2010).

Los primeros usos de la estadística implicaron la recopilación de datos y la elaboración de gráficas, para describir diversos aspectos de un estado o de un país. Los hogares, gobiernos y empresas se apoyan en estos datos estadísticos para dirigir sus acciones (Triola, 2009). Por ejemplo, en Ecuador se realizan censos poblacionales periódicos cada 10 años para obtener información sobre educación, salud, trabajo, empresas, tecnología, medio ambiente y servicios públicos; información utilizada por los líderes empresariales y gubernamentales para realizar planeaciones estratégicas.

Estadística descriptiva. También llamada estadística deductiva, pone en evidencia aspectos característicos de la variable estudiada que sirven para efectuar comparaciones o dependencias que pudieran existir en una serie de individuos sin pretender sacar conclusiones de tipo más general. En esta, la presentación de datos se lo realiza mediante tablas y gráficos estadísticos (Fernández, Cordero & Córdoba, 2002).

Estadística inferencial. También llamada estadística inductiva o analítica, implica la formulación de conclusiones o inferencias sobre un parámetro. Trata de dar explicación al comportamiento del conjunto de observaciones y las causas que lo originan para dar significancia o confiabilidad a los resultados (Agresti, Finlay, De Battisti y Porro, 2018). Tiene gran aplicación en el ámbito muestral, extrayendo así conclusiones más allá de la mera muestra estadística (Triola, 2018).

En la práctica, se utiliza para realizar pruebas de hipótesis, estimar parámetros desconocidos de una población, y hacer predicciones sobre futuros eventos. Algunas aplicaciones comunes incluyen el análisis de encuestas, la evaluación de resultados de experimentos, y la predicción de tendencias en los mercados financieros.

1.3 POBLACIÓN, MUESTRA Y SUS ELEMENTOS

Población. También conocido como universo, es el conjunto de todos los elementos o medidas que comparten una característica común y que son el motivo de nuestra investigación. (Dousdebés, 2021).

Dependiendo del número de unidades que tenga la población, esto se puede considerar como finita o infinita. Una población es finita cuando está compuesta de un número limitado de elementos, por ejemplo, el número de estudiantes universitarios en una región, el número de vehículos de una ciudad, etc. Y la población infinita es aquella que tiene una gran cantidad de componentes y aunque su totalidad no es infinita su gran tamaño hace muy difícil cuantificarla, por ejemplo, la cantidad especies que se encuentran en el mundo animal, la totalidad de árboles de mandarina en Ecuador, etc.

En un estudio científico la definición de la población es muy importante ya que afecta directamente la precisión y validez de los datos, y aunque lo ideal es investigar a la totalidad de elementos que conforman la población esto no siempre es posible debido a las limitaciones de tiempo, recursos o acceso a todos los elementos, en estos casos se puede seleccionar de entre ellos un grupo más pequeño al que llamaremos “muestra” cuyo tamaño se calcula de tal manera que las conclusiones obtenidas de esta sean válidas para la población.

Muestra. Se define como un conjunto de elementos seleccionados adecuadamente de una población en particular, es decir, parte de una población o población. Cuando se selecciona una muestra, se supone que el análisis realizado sobre esa muestra puede producir conclusiones similares a las que se obtendrían si se examinaran todos los elementos de la población (Posada, 2016).

Su tamaño puede ser calculado mediante fórmulas estadísticas o también definido por el mismo investigador cuando este tiene gran trayectoria y experiencia investigativa, no obstante, a nivel general se dice que esta no debe ser ni demasiado grande como para perder su objetivo el cual es disminuir los esfuerzos en el estudio de variables, ni tan pequeña como para perder representatividad. En cuanto a los elementos que conforman la muestra, estos deben ser seleccionados al azar.

El experimento. Es una herramienta fundamental en la investigación científica. Consiste en la manipulación de variables controladas para observar y medir los efectos de un tratamiento o intervención en un fenómeno o proceso específico. Los experimentos pueden ser realizados tanto en laboratorios como en entornos naturales, dependiendo de los objetivos de la investigación. Con relación a la población y muestra, se puede decir que en el experimento "*La población serán todos los datos que podríamos haber recopilado si repitiéramos el experimento una gran cantidad de veces (un número infinito de veces) en las mismas condiciones, mientras que la muestra serán los datos realmente recopilados por el experimento*" (Ramachandran, Tsokos, 2009).

Espacio muestral. Refiere al conjunto de todos los posibles resultados posibles de un experimento aleatorio (Amaral, 2015). Es decir, es el conjunto de todos los resultados que pueden ocurrir al realizar un experimento. El espacio muestral puede ser finito o infinito dependiendo del experimento, además, puede ser discreto o continuo, dependiendo de si los resultados son números enteros o reales.

Los elementos. Los elementos que integran la población o la muestra pueden corresponder a personas, objetos o cosas. Además, el elemento puede ser una entidad simple (una persona) o una entidad compleja (una familia), y se denomina unidad investigada. Es importante resaltar el hecho de que, a pesar de encontrarse una población constituida por un grupo de elementos, a la estadística no le interesa el elemento en sí, sino sus características (Martínez, 2020).

Marco de muestreo. Conocida también como marco de referencia, es una lista, mapa o registro de los individuos u objetos que tienen que ser muestreados, o a su vez, los materiales o dispositivos utilizados para acceder a los miembros de una población de interés (Luzardo y Jimenez, 2018) (Devore, 2008).

El marco de muestreo se utiliza para garantizar que la muestra seleccionada sea representativa de la población objetivo y que los resultados obtenidos puedan generalizarse a la población en su conjunto.

Para crear un marco de muestreo, es necesario identificar todos los elementos de la población objetivo y asignarles un número o identificador único, a continuación, se selecciona una muestra aleatoria de estos elementos.

Características (o caracteres). Es una propiedad o atributo de la población o muestra que permite medirlos o contarlos, y se utilizan para describir y analizar los datos estadísticos. Algunos caracteres son mensurables y se describen numéricamente, por tal motivo se denominan caracteres cuantitativos o variables, por ejemplo: la estatura, peso, ingreso, valor, producción, ventas, etc., mientras que otros se expresan mediante palabras por no ser mensurables, pero si cuantificables, por tal motivo se denominan caracteres cualitativos o atributos, por ejemplo: la profesión de un trabajador, la calidad de un producto, etc. (Martínez, 2020).

Estadísticas temporales. Llamadas también series de tiempo o series cronológicas son aquellas mediciones u observaciones realizadas de forma periódica, por ejemplo. Cada 10 minutos, cada 2 horas, diariamente, semanalmente, mensual, trimestral, anual, etc. en caso contrario, cuando un muestreo se los realiza de forma aislada se dice que estas mediciones son atemporales (Martínez, 2020).

La importancia de la estadística temporal radica en el análisis de variación de los datos a lo largo del tiempo para detectar patrones y tendencias, mientras que la estadística atemporal permite detectar relaciones y correlaciones entre variables.

Parámetro. Es aquella característica numérica desconocida de la distribución de los elementos de la población (De Souza, Pinheiro y Bastos, 2018). Es decir, es una medida numérica que se utiliza para describir una población. Por ejemplo. la media, la varianza y la desviación estándar son parámetros comunes utilizados para describir una población normal. Como se podrá deducir, una población puede tener varias características y por tanto varios parámetros.

Estimador. Es la función de la muestra, construida con el propósito de representar o estimar un parámetro de interés en la población (De Souza, Pinheiro y Bastos, 2018). Dicho de otra forma, esta es una regla matemática que se utiliza para aproximar o estimar un valor desconocido de una población a partir de una muestra de datos y se utiliza para hacer inferencias acerca de la población a partir de la información disponible. Por ejemplo, si queremos conocer el promedio de edad de una población, pero no podemos medir la edad de todas las personas, podemos tomar una muestra aleatoria y utilizar un estimador para aproximar este valor.

1.4 FUENTES DE INFORMACIÓN ESTADÍSTICA.

La información estadística según la fuente de origen puede clasificarse de dos formas:

Fuentes primarias. También llamadas fuentes de datos internas, refiere a los registros almacenados internamente en la organización o información recogida de primera mano por el (los) sujeto (s) que realiza (n) la investigación a través de cualquier método como entrevistas, encuestas, experimentos, observaciones, etc.

Fuentes secundarias. También llamadas fuentes de datos externas son aquellos datos producidos por organizaciones o investigadores externos, estos generalmente se publican en páginas web, libros, revistas científicas, informes gubernamentales, etc. En este apartado es importante recordar que las buenas prácticas éticas de investigación obligan al investigador a verificar la veracidad de las fuentes elegidas, debido a que, la incorporación de datos erróneos pudiese alterar significa mente los resultados de nuestra investigación.

1.5 LA INVESTIGACIÓN ESTADÍSTICA

El proceso de la investigación estadística es una actividad que requiere tomar atención en varios aspectos exploratorios que producirán muchas funciones, cuyos resultados dependen en gran medida de los fines que se persigan, la naturaleza de los fenómenos estudiados y complejidad de los fenómenos estudiados (Martínez, 2020). En este sentido, el proceso de la investigación científica supone varias la ejecución de varias fases:

Planteamiento del problema. Consiste en definir los objetivos, la hipótesis y las variables de la investigación, precisando la población objetivo y los límites del estudio.

Planeación de actividades. Permite describir el rol de cada investigador y las actividades que tendrá a cargo, así como también define los instrumentos de recolección de datos y el procedimiento de muestreo aplicar.

Recolección de datos. Consiste en la aplicación de la técnica de muestreo planeada, la depuración de la información para eliminar posibles datos anómalos que pueden incluir errores de observación o transcripción.

*Análisis de descriptivo de la información. E*n esta fase se construyen tablas estadísticas a fin de resumir la información obtenida para su posterior procesamiento, también se realiza la presentación de datos a través de gráficos estadísticos que otorguen un entendimiento general sobre el comportamiento de las variables estudiadas.

Análisis inferencial de los datos. En esta fase se aplican técnicas estadísticas específicas como pruebas de hipótesis, intervalos de confianza y análisis de regresión para hallar conclusión a partir de hechos, proposiciones o principios a partir de la población o muestra.

Elaboración de conclusiones. Finalmente, el investigador debe redactar un documento formal con los resultados obtenidos, estos pueden ser informes, artículos científicos, libros.

UNIDAD DOS
DISTRIBUCIÓN DE FRECUENCIAS

2. INTRODUCCIÓN

Una de las primeras estrategias usadas para es el análisis de la información recogida es la distribución de frecuencias, para ello debemos organizarla y agruparla de modo que pueda ser más comprensible y operable. Al analizar esta información se pueden identificar patrones y tendencias en los datos, lo que puede ayudar a los investigadores e instituciones a tomar decisiones informadas. Por ejemplo, si una empresa observa una distribución de frecuencia sesgada hacia valores más altos de una variable, puede indicar que hay un problema en su proceso de producción o en su modelo de costos.

2.1 DISTRIBUCIÓN DE FRECUENCIAS

La distribución de frecuencias es una herramienta estadística que permite agrupar o resumir grandes cantidades de información en tablas de datos distribuidas en filas ordenadas. Esta es una forma de visualizar la distribución de los datos de modo que se puedan identificar patrones y tendencias. Existen dos formas de distribuir la frecuencia de los datos.

Distribución de frecuencia simples. Esta distribución muestra la frecuencia con la que ocurre cada uno de los valores de la muestra (sin eliminar ninguno). Este método es útil para grupos de datos donde no existen variaciones significativas entre sus valores, por ejemplo, si se recogen las edades de 485 estudiantes de bachillerato general unificado en Ecuador, independientemente de que tan grande sea la muestra los valores apuntados posiblemente estarán entre 15 a 18 años, lo cual es un rango muy pequeño y fácilmente se pudiera construir una tabla con filas para cada dato, como se muestra a continuación.

Tabla 1. Edades de estudiantes de bachillerato

Edad (años)	Número de estudiantes
15	131
16	149
17	112
18	93

Distribución de frecuencia por intervalos. A diferencia de la distribución simple, la distribución por intervalos se usa cuando los datos recogidos en la muestra son muy variados entre ellos. En este caso conviene formar grupos de datos (llamados clases) para apuntarlos en cada fila de la tabla, por ejemplo, del mismo grupo de estudiantes del ejemplo anterior ahora medimos su peso en Kilogramos, en este caso, seguramente los valores encontrados pudieran estar comprendidos entre 52.6 a 59,8 que son los pesos promedios de jóvenes entre 15 y 18 años en Ecuador (Tarupi, Lepage, Felix, Monnier, Hauspie, Roelants, Hidalgo y Vercauteren, 2020). Una posible distribución pudiera ser la siguiente.

Tabla 2. Peso de estudiantes de bachillerato

Peso (Kg)			Número de estudiantes
52,6	a	53,7	69
53,8	a	54,9	87
55	a	56,1	76
56,2	a	57,3	85
57,4	a	58,5	92
58,6	a	59,7	76

2.2 TABLAS DE FRECUENCIAS

Se trata de una tabla que muestra la frecuencia con la que ocurre cada valor o rango de valores de un conjunto de datos obtenidos de una población o una muestra. Describe en forma resumida de la frecuencia de cada una de las distribuciones encontradas.

2.3 TABLA DE FRECUENCIAS PARA UNA DISTRIBUCIÓN SIMPLE

El procedimiento para su construcción es la siguiente.

La tabla de frecuencias contiene las siguientes columnas.

Tabla 3. Enunciados de una tabla de frecuencias simple

X $(Marca\ de\ clase)$	Fa $(Frecuencia\ absoluta)$	faa $\left(\begin{matrix}Frecuencia\ absoluta\\ acumulada\end{matrix}\right)$	fr $(Frecuencia\ relativa)$	fra $\left(\begin{matrix}Frecuencia\ relativa\\ acumulada\end{matrix}\right)$	Fra% $\left(\begin{matrix}Frecuencia\ relativa\ a\\ cumulada\ porcentual\end{matrix}\right)$

Para resolver el ejemplo de la tabla 1. se procede a insertar la edad en años en la columna de la marca de clase y el número de estudiantes en la columna de la frecuencia absoluta de la siguiente manera.

Tabla 4. Agrupamiento de datos en una tabla de frecuencias simple

X $\left(\begin{matrix}Marca\ de\\ clase\end{matrix}\right)$	Fa $\left(\begin{matrix}Frecuencia\\ absoluta\end{matrix}\right)$	faa $\left(\begin{matrix}Frecuencia\\ absoluta\\ acumulada\end{matrix}\right)$	fr $\left(\begin{matrix}Frecuencia\\ relativa\end{matrix}\right)$	fra $\left(\begin{matrix}Frecuencia\\ relativa\\ acumulada\end{matrix}\right)$	Fra% $(fra\ porcentual)$
15	131				
16	149				
17	112				
18	93				

La columna de frecuencia absoluta acumulada es la suma de las frecuencias absolutas anteriores incluida la de su fila. Por ejemplo, la **faa** de la segunda fila (280) es el resultado de sumar la fa de la segunda (149) y primera fila (131), la faa de la tercera fila (392) es el resultado de sumar las **fa** de la tercera (112), segunda (149) y primera fila (131), etc.

Note que la frecuencia absoluta acumulada de la última fila (485) es igual al N (número total de datos de la muestra).

Tabla 5. Faa en una tabla de frecuencias simple

X $\binom{Marca\ de}{clase}$	fa $\binom{Frecuencia}{absoluta}$	faa $\left(\begin{matrix}Frecuencia\\absoluta\\acumulada\end{matrix}\right)$	fr $\binom{Frecuencia}{relativa}$	fra $\left(\begin{matrix}Frecuencia\\relativa\\acumulada\end{matrix}\right)$	Fra% $(fra\ porcentual)$
15	131	131			
16	149	280			
17	112	392			
18	93	485			

La columna de la frecuencia relativa se calcula dividiendo la frecuencia absoluta de su propia fila para el total de datos, y se apunta redondeando a cuatro cifras decimales. Es decir.

$$fr = \frac{fa}{N}$$

Donde N es la cantidad de datos de la muestra.

Por ejemplo, la fr de la primera fila (0,2701) aparece de dividir la fa de la primera fila (131) para N=485, la fr de la segunda fila (0,3072) aparece de dividir la fa de la segunda fila (149) para N=485, etc.

Tabla 6. Fr en una tabla de frecuencias simple

X $\left(\begin{matrix}Marca\ de\\ clase\end{matrix}\right)$	fa $\left(\begin{matrix}Frecuencia\\ absoluta\end{matrix}\right)$	faa $\left(\begin{matrix}Frecuencia\\ absoluta\\ acumulada\end{matrix}\right)$	fr $\left(\begin{matrix}Frecuencia\\ relativa\end{matrix}\right)$	fra $\left(\begin{matrix}Frecuencia\\ relativa\\ acumulada\end{matrix}\right)$	Fra% $(fra\ porcentual)$
15	131	131	0,2701		
16	149	280	0,3072		
17	112	392	0,2309		
18	93	485	0,1918		

La columna de la frecuencia relativa acumulada, al igual que la faa, aparece de la sumar las fr anteriores incluida la de su propia fila. Por ejemplo, la fra de la segunda fila (0,5773) aparece de la sumar $0{,}3072 + 0{,}2701$, la fra de la tercera fila (0,8082) aparece de la sumar $0{,}2309 + 0{,}3072 + 0{,}2701$, etc.

Tabla 7. Fra en una tabla de frecuencias simple

X $\left(\begin{matrix}Marca\ de\\ clase\end{matrix}\right)$	fa $\left(\begin{matrix}Frecuencia\\ absoluta\end{matrix}\right)$	faa $\left(\begin{matrix}Frecuencia\\ absoluta\\ acumulada\end{matrix}\right)$	fr $\left(\begin{matrix}Frecuencia\\ relativa\end{matrix}\right)$	fra $\left(\begin{matrix}Frecuencia\\ relativa\\ acumulada\end{matrix}\right)$	Fra% $(fra\ porcentual)$
15	131	131	0,2701	0,2701	
16	149	280	0,3072	0,5773	
17	112	392	0,2309	0,8082	
18	93	485	0,1918	1	

Finalmente, la columna de frecuencia relativa acumulada porcentual se calcula transformando en porcentaje la fra. De esta forma.

$$fra\% = Fra * 100$$

La tabla de frecuencias simples terminada quedaría de la siguiente forma.

Tabla 8. Tabla de frecuencias simple

X $\left(\begin{matrix}Marca\ de\\ clase\end{matrix}\right)$	fa $\left(\begin{matrix}Frecuencia\\ absoluta\end{matrix}\right)$	faa $\left(\begin{matrix}Frecuencia\\ absoluta\\ acumulada\end{matrix}\right)$	fr $\left(\begin{matrix}Frecuencia\\ relativa\end{matrix}\right)$	fra $\left(\begin{matrix}Frecuencia\\ relativa\\ acumulada\end{matrix}\right)$	Fra% $(fra\ porcentual)$
15	131	131	0,2701	0,2701	27,01%
16	149	280	0,3072	0,5773	57,73%
17	112	392	0,2309	0,8082	80,82%
18	93	485	0,1918	1	100,00%

2.4 TABLA DE FRECUENCIAS PARA UNA DISTRIBUCIÓN POR INTERVALOS

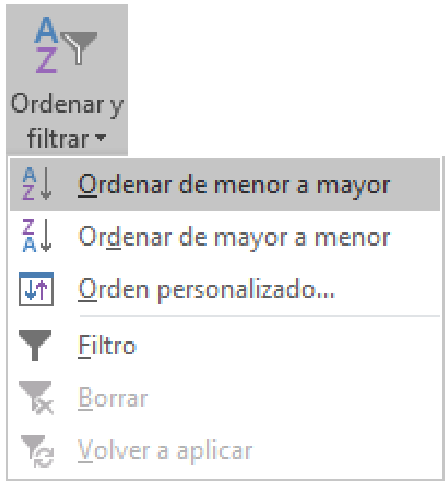
Este tipo de tablas son más elaboradas que las del apartado anterior. El método de construcción presentado en esta obra, al cual hemos llamado “Método de Freire”, está fundamentado en métodos tradiciones, el cuál fue aplicado en los cursos de estadística de una institución de educación superior en Ecuador por un tiempo de prueba de siete años. Las operaciones matemáticas utilizadas tienen armonía con las funciones de Microsoft Excel 2013 en adelante, a fin de que el método pueda ser fácilmente automatizado. El proceso se describe a continuación.

En este caso, la tabla de frecuencias está compuesta por las siguientes columnas.

Tabla 9. Enunciados de una tabla de frecuencias por intervalos

Xri $\left(\begin{matrix}Limite\ real\\ inferior\end{matrix}\right)$	Xi $\left(\begin{matrix}Limite\\ inferior\end{matrix}\right)$	Xs $\left(\begin{matrix}Limite\\ superior\end{matrix}\right)$	Xrs $\left(\begin{matrix}Limite\ real\\ superior\end{matrix}\right)$	X $\left(\begin{matrix}Marca\ de\\ clase\end{matrix}\right)$	fa $\left(\begin{matrix}Frecuencia\\ absoluta\end{matrix}\right)$	faa $\left(\begin{matrix}Frecuencia\ absoluta\\ acumulada\end{matrix}\right)$	fr $\left(\begin{matrix}Frecuencia\\ relativa\end{matrix}\right)$	fra $\left(\begin{matrix}Frecuencia\ relativa\\ acumulada\end{matrix}\right)$	Fra% $\left(\begin{matrix}Frecuencia\ relativa\ a\\ cumulada\ porcentual\end{matrix}\right)$

El primer paso es, *ordenar los datos de la muestra de menor a mayor*, si está trabajando en Microsoft Excel 2021 (ME2021) puede usar la función "Ordenar de menor a mayor" ubicada en la pestaña Inicio.



Caso 1.

Parte de un estudio de conocimiento, se evaluó a noventa jóvenes bachilleres en el área de matemática superior. Las calificaciones obtenidas (sobre 10 puntos) se muestran a continuación.

4,6	5,1	2,3	3,1	3,7	2,0	2,8	3,4	4,0	3,3	3,8	3,0	3,7	6,1	7,0	8,4	5,0	9,4
6,8	3,8	4,3	2,4	3,2	3,8	4,7	8,0	9,4	4,7	5,2	3,7	4,2	5,0	4,5	5,1	3,4	4,0
6,2	7,2	9,2	6,0	6,8	4,9	5,3	6,0	4,3	4,3	5,0	5,5	5,9	6,5	7,6	5,9	6,4	7,5
7,1	8,8	5,8	6,2	7,2	2,5	3,3	8,1	5,2	2,6	3,4	3,8	2,2	2,0	2,8	3,6	4,1	5,6
6,2	5,0	5,5	2,5	6,6	5,4	5,4	6,0	7,7	5,8	6,3	7,4	2,0	2,7	4,2	5,0	2,1	2,9

Los datos ordenados del ejemplo 1 son los siguientes.

Dato Menor →	**2,0**	2,7	3,4	4,0	4,7	5,2	5,9	6,4	7,5	
	2,0	2,8	3,4	4,0	4,7	5,2	5,9	6,5	7,6	
	2,0	2,8	3,6	4,1	4,9	5,3	6,0	6,6	7,7	
	2,1	2,9	3,7	4,2	5,0	5,4	6,0	6,8	8,0	
	2,2	3,0	3,7	4,2	5,0	5,4	6,0	6,8	8,1	
	2,3	3,1	3,7	4,3	5,0	5,5	6,1	7,0	8,4	
	2,4	3,2	3,8	4,3	5,0	5,5	6,2	7,1	8,8	
	2,5	3,3	3,8	4,3	5,0	5,6	6,2	7,2	9,2	
	2,5	3,3	3,8	4,5	5,1	5,8	6,2	7,2	9,4	
	2,6	3,4	3,8	4,6	5,1	5,8	6,3	7,4	**9,4**	← Dato Mayor

El segundo paso es, *calcular el Número de intervalos de clase (NIC),* para determinar la cantidad de grupos (o filas) que tendrá la tabla de frecuencias. En algunos casos, cuando el investigador tiene amplia trayectoria y experiencia en el campo especifico del estudio, el NIC es asignado a priori, considerando simplemente el hecho de que tomar un número demasiado pequeño de tramos hará que la tabla obtenida sea poco representativa, por el contrario, asignar un valor demasiado grande hará que pierda el objetivo construir una tabla de frecuencia, el cual es disminuir el tiempo de procesamiento de los datos. Recuerde que, a cada grupo de información se le denomina clase.

Por cuestiones didácticas se usará la regla de Sturges para el cálculo de número de intervalos, la fórmula es la siguiente.

$$NIC = 1 + \log_2 N$$

Donde N es la cantidad de datos de la muestra.

Si usa calculadora de mano y no cuenta con la función de log en base N, la función anterior se podría escribir de la siguiente manera.

$$NIC = 1 + \frac{\log N}{\log 2}$$

La respuesta obtenida siempre se sube al "número entero inmediato superior".

El número de intervalos de clase para el ejemplo 1 es la siguiente

$$NIC = 1 + \log_2 90 = 7{,}4918 \approx 8$$

Notas Importantes

- *El NIC siempre es un número entero. Pero si la respuesta de Sturges da como resultado un número decimal, este se sube al número entero inmediato superior, en nuestro ejercicio de 7,4918 subió a 8.*
- *El procedimiento de subir al inmediato superior indica que no importa si la respuesta hubiera sido 7,0001; 7,5 o 7,9999, el NIC siempre subiría a 8. (no confundir este procedimiento con el redondeo de datos, son procesos diferentes)*
- *En ME19, la función para calcular log en cualquier base es "$= LOG(dato;\ base)$", y la función para subir al inmediato superior es "=REDONDEAR.MAS(dato; num_decimales)"*

El tercer paso consiste en *calcular el ancho de intervalo de clase (AIC)*, esto permite conocer el ancho que tendrá cada grupo de datos de las clases. La función es la siguiente.

$$AIC = \frac{Dato\ mayor\ de\ la\ muestra - Dato\ menor\ de\ la\ muestra}{NIC}$$

El resultado del AIC deberá tener la misma cantidad de cifras decimales que la muestra, para ello, las cifras que están demás se bajan al inmediato inferior.

El ancho de intervalo de clase para el ejemplo 1 es

$$AIC = \frac{9{,}4 - 2{,}0}{8} = 0{,}925\ \approx 0{,}9$$

Notas importantes:

- *El valor del AIC debe contener la misma cantidad de dígitos decimales que los datos de la muestra, en nuestro caso observamos que los números que conforman la muestra tiene un digito decimal. Pero en nuestro ejemplo el AIC da como resultado 0,925 que es un número de tres cifras decimales, por tanto, se deja en 0,9 despreciando las dos cifras decimales que están demás.*
- *Este procedimiento bajar al inmediato inferior significa que no importa si la respuesta hubiera sido 0,901; 0,950; 0,999; siempre quedara en 0,9 que es inmediato inferior de una cifra decimal. (no confundir este procedimiento con el redondeo de datos, son procesos diferentes).*
- *No olvide trabajar siempre según el número de cifras decimales de la muestra, por ejemplo, si los datos de la muestra son números enteros el AIC también deberá tener cero decimales, y si los datos de la muestra son números con dos cifras decimales el AIC también se deberá tener dos cifras decimales.*
- *En ME19, la función para bajar al inmediato inferior es "=REDONDEAR.MENOS(dato; num_decimales)"*

El cuarto paso consiste en *construir la tabla de frecuencias*, para ello seguimos las siguientes instrucciones.

En el Xi de la primera clase se coloca el dato menor de la muestra.

Tabla 10. Xi 1ra clase, en una tabla de frecuencias por intervalos

Núm. de clase	Xri (Limite real inferior)	Xi (Limite inferior)	Xs (Limite superior)	Xrs (Limite real superior)	X (Marca de clase)	fa (Frecuencia absoluta)	faa (Frecuencia absoluta acumulada)	fr (Frecuencia relativa)	fra (Frecuencia relativa acumulada)	Fra% (fra porcentual)
1		2,0								
2										
3										
4										
5										
6										
7										
8										

El Xs de la primera clase (2,9), se calcula sumando el Xi de la primera clase (2,0) más el AIC (0,9).

Tabla 11. Xs 1ra clase, en una tabla de frecuencias por intervalos

+AIC

Núm. de clase	Xri (Limite real inferior)	Xi (Limite inferior)	Xs (Limite superior)	Xrs (Limite real superior)	X (Marca de clase)	fa (Frecuencia absoluta)	faa (Frecuencia absoluta acumulada)	fr (Frecuencia relativa)	fra (Frecuencia relativa acumulada)	Fra% (fra porcentual)
1		2,0	2,9							
2										
3										
4										
5										
6										
7										
8										

El Xi de la segunda clase (3,0), es el número de una cifra decimal que le sigue al Xs de la primera clase (2,9).

No olvide trabajar siempre según el número de cifras decimales de la muestra, en este caso la muestra contiene números de una cifra decimal y por ese motivo Xi y Xs también tienen una cifra decimal, si la muestra estuviera conformada por números enteros el Xi y Xs serían números enteros, si la muestra estuviera conformada por números de dos cifras decimales el Xi y Xs también serían números de dos cifras decimales, etc.

Tabla 12. Xi 2da clase, en una tabla de frecuencias por intervalos

Núm. de clase	Xri (Limite real inferior)	Xi (Limite inferior)	Xs (Limite superior)	Xrs (Limite real superior)	X (Marca de clase)	fa (Frecuencia absoluta)	faa (Frecuencia absoluta acumulada)	fr (Frecuencia relativa)	fra (Frecuencia relativa acumulada)	Fra% (fra porcentual)
1		2,0	2,9							
2		3,0								
3										
4										
5										
6										
7										
8										

Repetimos las tres instrucciones anteriores hasta completar la octava clase. Es decir, el Xs de la segunda clase (3,9) aparece de sumar el Xi de la segunda clase (3,0) más el AIC (0,9), luego, el Xi de la tercera clase (4,0) es el número de una cifra decimal que le sigue al Xs de la segunda clase (3,9), etc.

Tabla 13. Xi y Xs en una tabla de frecuencias por intervalos

+AIC

Núm. de clase	Xri (Limite real inferior)	Xi (Limite inferior)	Xs (Limite superior)	Xrs (Limite real superior)	X (Marca de clase)	fa (Frecuencia absoluta)	faa (Frecuencia absoluta acumulada)	fr (Frecuencia relativa)	fra (Frecuencia relativa acumulada)	Fra% (fra porcentual)
1		2,0	2,9							
2		3,0	3,9							
3		4,0	4,9							
4		5,0	5,9							
5		6,0	6,9							
6		7,0	7,9							
7		8,0	8,9							
8		9,0	9,9							

El límite real superior de la primera clase (2,95), se calcula promediando el Xs de la primera clase (2,9) con el Xi de la segunda clase (3,0).

Tabla 14. Xrs 1ra clase, en una tabla de frecuencias por intervalos

Núm. de clase	Xri (Limite real inferior)	Xi (Limite inferior)	Xs (Limite superior)	Xrs (Limite real superior)	X (Marca de clase)	fa (Frecuencia absoluta)	faa (Frecuencia absoluta acumulada)	fr (Frecuencia relativa)	fra (Frecuencia relativa acumulada)	Fra% (fra porcentual)
1		2,0	2,9	2,95						
2		3,0	3,9							
3		4,0	4,9							
4		5,0	5,9							
5		6,0	6,9							
6		7,0	7,9							
7		8,0	8,9							
8		9,0	9,9							

El Ancho real (AR) (0,05), se calcula hallando la diferencia entre Xrs (2,95) y Xs (2,9) de la primera clase.

Tabla 15. Ancho real en una tabla de frecuencias por intervalos

AR=0,05

Núm. de clase	Xri (Límite real inferior)	Xi (Límite inferior)	Xs (Límite superior)	Xrs (Límite real superior)	X (Marca de clase)	fa (Frecuencia absoluta)	faa (Frecuencia absoluta acumulada)	fr (Frecuencia relativa)	fra (Frecuencia relativa acumulada)	Fra% (fra porcentual)
1		2,0	2,9	2,95						
2		3,0	3,9							
3		4,0	4,9							
4		5,0	5,9							
5		6,0	6,9							
6		7,0	7,9							
7		8,0	8,9							
8		9,0	9,9							

El Xrs de la segunda a la octava clase, se calcula sumando el Xs de cada clase con el AR.

Tabla 16. Xrs en una tabla de frecuencias por intervalos

+AR

Núm. de clase	Xri (Límite real inferior)	Xi (Límite inferior)	Xs (Límite superior)	Xrs (Límite real superior)	X (Marca de clase)	Fa (Frecuencia absoluta)	faa (Frecuencia absoluta acumulada)	fr (Frecuencia relativa)	fra (Frecuencia relativa acumulada)	Fra% (fra porcentual)
1		2,0	2,9	2,95						
2		3,0	3,9	3,95						
3		4,0	4,9	4,95						
4		5,0	5,9	5,95						
5		6,0	6,9	6,95						
6		7,0	7,9	7,95						
7		8,0	8,9	8,95						
8		9,0	9,9	9,95						

El límite real inferior, se calcula restando el Xi de cada clase con el AR.

Tabla 17. Xri en una tabla de frecuencias por intervalos

- AR

Núm. de clase	Xri (Limite real inferior)	Xi (Limite inferior)	Xs (Limite superior)	Xrs (Limite real superior)	X (Marca de clase)	Fa (Frecuencia absoluta)	faa (Frecuencia absoluta acumulada)	fr (Frecuencia relativa)	fra (Frecuencia relativa acumulada)	Fra% (fra porcentual)
1	1,95	2,0	2,9	2,95						
2	2,95	3,0	3,9	3,95						
3	3,95	4,0	4,9	4,95						
4	4,95	5,0	5,9	5,95						
5	5,95	6,0	6,9	6,95						
6	6,95	7, 0	7,9	7,95						
7	7,95	8, 0	8,9	8,95						
8	8,95	9,0	9,9	9,95						

La marca de clase de la primera clase (2,45), se calcula promediando el Xi (2,0) y el Xs (2,9) de la primera clase. Es decir:

$$X_{(1ra\ clase)} = \frac{Xi_{(1ra\ clase)} + Xs_{(1ra\ clase)}}{2} = \frac{2{,}0 + 2{,}9}{2} = 2{,}45$$

Este valor se apunta con todas las cifras decimales que tuviera el resultado.

Tabla 18. X 1ra clase, en una tabla de frecuencias por intervalos

Núm. de clase	Xri (Limite real inferior)	Xi (Limite inferior)	Xs (Limite superior)	Xrs (Limite real superior)	X (Marca de clase)	fa (Frecuencia absoluta)	faa (Frecuencia absoluta acumulada)	fr (Frecuencia relativa)	fra (Frecuencia relativa acumulada)	Fra% (fra porcentual)
1	1,95	2,0	2,9	2,95	2,45					
2	2,95	3,0	3,9	3,95						
3	3,95	4,0	4,9	4,95						
4	4,95	5,0	5,9	5,95						
5	5,95	6,0	6,9	6,95						
6	6,95	7, 0	7,9	7,95						
7	7,95	8, 0	8,9	8,95						
8	8,95	9,0	9,9	9,95						

Las marcas de clase de las demás clases se calculan de la misma manera, promediando sus respectivos Xi y Xs.

Tabla 19. X en una tabla de frecuencias por intervalos

Núm. de clase	Xri (Limite real inferior)	Xi (Limite inferior)	Xs (Limite superior)	Xrs (Limite real superior)	X (Marca de clase)	fa (Frecuencia absoluta)	faa (Frecuencia absoluta acumulada)	fr (Frecuencia relativa)	fra (Frecuencia relativa acumulada)	Fra% (fra porcentual)
1	1,95	2,0	2,9	2,95	2,45					
2	2,95	3,0	3,9	3,95	3,45					
3	3,95	4,0	4,9	4,95	4,45					
4	4,95	5,0	5,9	5,95	5,45					
5	5,95	6,0	6,9	6,95	6,45					
6	6,95	7, 0	7,9	7,95	7,45					
7	7,95	8, 0	8,9	8,95	6,45					
8	8,95	9,0	9,9	9,95	7,45					

La frecuencia absoluta de la primera clase (14), se calcula contando la cantidad de números de la muestra que están en el rango de Xi (2,0) a Xs (2,9) de la primera clase.

2,0	2,7	3,4	4,0	4,7	5,2	5,9	6,4	7,5
2,0	2,8	3,4	4,0	4,7	5,2	5,9	6,5	7,6
2,0	2,8	3,6	4,1	4,9	5,3	6,0	6,6	7,7
2,1	2,9	3,7	4,2	5,0	5,4	6,0	6,8	8,0
2,2	3,0	3,7	4,2	5,0	5,4	6,0	6,8	8,1
2,3	3,1	3,7	4,3	5,0	5,5	6,1	7,0	8,4
2,4	3,2	3,8	4,3	5,0	5,5	6,2	7,1	8,8
2,5	3,3	3,8	4,3	5,0	5,6	6,2	7,2	9,2
2,5	3,3	3,8	4,5	5,1	5,8	6,2	7,2	9,4
2,6	3,4	3,8	4,6	5,1	5,8	6,3	7,4	9,4

Tabla 20. fa 1ra clase, en una tabla de frecuencias por intervalos

Núm. de clase	Xri (Limite real inferior)	Xi (Limite inferior)	Xs (Limite superior)	Xrs (Limite real superior)	X (Marca de clase)	fa (Frecuencia absoluta)	faa (Frecuencia absoluta acumulada)	fr (Frecuencia relativa)	fra (Frecuencia relativa acumulada)	Fra% (fra porcentual)
1	1,95	2,0	2,9	2,95	2,45	14				
2	2,95	3,0	3,9	3,95	3,45					
3	3,95	4,0	4,9	4,95	4,45					
4	4,95	5,0	5,9	5,95	5,45					
5	5,95	6,0	6,9	6,95	6,45					
6	6,95	7, 0	7,9	7,95	7,45					
7	7,95	8, 0	8,9	8,95	6,45					
8	8,95	9,0	9,9	9,95	7,45					

Repetimos el proceso para las demás marcas de clase.

2,0	2,7	3,4	4,0	4,7	5,2	5,9	6,4	7,5
2,0	2,8	3,4	4,0	4,7	5,2	5,9	6,5	7,6
2,0	2,8	3,6	4,1	4,9	5,3	6,0	6,6	7,7
2,1	2,9	3,7	4,2	5,0	5,4	6,0	6,8	8,0
2,2	3,0	3,7	4,2	5,0	5,4	6,0	6,8	8,1
2,3	3,1	3,7	4,3	5,0	5,5	6,1	7,0	8,4
2,4	3,2	3,8	4,3	5,0	5,5	6,2	7,1	8,8
2,5	3,3	3,8	4,3	5,0	5,6	6,2	7,2	9,2
2,5	3,3	3,8	4,5	5,1	5,8	6,2	7,2	9,4
2,6	3,4	3,8	4,6	5,1	5,8	6,3	7,4	9,4

Tabla 21. fa en una tabla de frecuencias por intervalos

Núm. de clase	Xri (Limite real inferior)	Xi (Limite inferior)	Xs (Limite superior)	Xrs (Limite real superior)	X (Marca de clase)	fa (Frecuencia absoluta)	faa (Frecuencia absoluta acumulada)	fr (Frecuencia relativa)	fra (Frecuencia relativa acumulada)	Fra% (fra porcentual)
1	1,95	2,0	2,9	2,95	2,45	14				
2	2,95	3,0	3,9	3,95	3,45	16				
3	3,95	4,0	4,9	4,95	4,45	13				
4	4,95	5,0	5,9	5,95	5,45	19				
5	5,95	6,0	6,9	6,95	6,45	13				
6	6,95	7, 0	7,9	7,95	7,45	8				
7	7,95	8, 0	8,9	8,95	6,45	4				
8	8,95	9,0	9,9	9,95	7,45	3				

La frecuencia absoluta acumulada, frecuencia relativa, relativa acumulada y porcentual se calculan de la misma forma como indico en la sección 2.4.

La tabla de frecuencia para distribución por intervalos quedaría de la siguiente forma.

Tabla 22. faa, fr, fra, fra% en una tabla de frecuencias por intervalos

Núm. de clase	Xri (Limite real inferior)	Xi (Limite inferior)	Xs (Limite superior)	Xrs (Limite real superior)	X (Marca de clase)	fa (Frecuencia absoluta)	faa (Frecuencia absoluta acumulada)	fr (Frecuencia relativa)	fra (Frecuencia relativa acumulada)	Fra% (fra porcentual)
1	1,95	2,0	2,9	2,95	2,45	14	14	0,1556	0,1556	15,56%
2	2,95	3,0	3,9	3,95	3,45	16	30	0,1778	0,3334	33,34%
3	3,95	4,0	4,9	4,95	4,45	13	43	0,1444	0,4778	47,78%
4	4,95	5,0	5,9	5,95	5,45	19	62	0,2111	0,6889	68,89%
5	5,95	6,0	6,9	6,95	6,45	13	75	0,1444	0,8333	83,33%
6	6,95	7, 0	7,9	7,95	7,45	8	83	0,0889	0,9222	92,22%
7	7,95	8, 0	8,9	8,95	6,45	4	87	0,0444	0,9666	96,66%
8	8,95	9,0	9,9	9,95	7,45	3	90	0,0333	0,9999	99,99%

A continuación, se realizará un segundo ejemplo de tabla de frecuencias por intervalos para una muestra con números enteros.

Caso 2.

Un estudio macroeconómico investiga las ventas anuales de 120 pequeñas empresas. Los valores en miles de dólares se muestran a continuación.

500	125	346	854	506	698	470	634	918	713	456	256	545	567	390	517	304	311
158	319	581	197	833	787	845	484	628	854	596	331	172	729	659	220	614	401
258	927	774	311	911	283	485	197	865	822	709	445	447	543	591	746	369	224
665	854	406	480	359	798	481	782	358	799	227	845	506	950	198	261	677	809
135	442	327	686	480	519	760	318	621	124	308	557	907	588	730	219	824	456
549	238	163	411	837	549	186	225	324	806	384	156	310	239	545	817	418	113
157	724	311	390	724	848	880	523	682	221	533	668						

Datos ordenados.

113	197	239	311	384	456	506	557	659	724	806	854
124	197	256	318	390	456	517	567	665	729	809	854
125	198	258	319	390	470	519	581	668	730	817	854
135	219	261	324	401	480	523	588	677	746	822	865
156	220	283	327	406	480	533	591	682	760	824	880
157	221	304	331	411	481	543	596	686	774	833	907
158	224	308	346	418	484	545	614	698	782	837	911
163	225	310	358	442	485	545	621	709	787	845	918
172	227	311	359	445	500	549	628	713	798	845	927
186	238	311	369	447	506	549	634	724	799	848	950

Número de intervalos de clase

$$NIC = 1 + \log_2 120 = 7{,}9069 \approx 8$$

Ancho de intervalo de clase

$$AIC = \frac{950-113}{8} = 104{,}625 \approx 104$$

Tabla de frecuencias

Tabla 23. Tabla de frecuencias por intervalos para el caso 2

Núm. de clase	Xri $\left(\mathit{Limite\ real\ inferior}\right)$	Xi $\left(\mathit{Limite\ inferior}\right)$	Xs $\left(\mathit{Limite\ superior}\right)$	Xrs $\left(\mathit{Limite\ real\ superior}\right)$	X $\left(\mathit{Marca\ de\ clase}\right)$	fa $\left(\mathit{Frecuencia\ absoluta}\right)$	faa $\left(\mathit{Frecuencia\ absoluta\ acumulada}\right)$	fr $\left(\mathit{Frecuencia\ relativa}\right)$	fra $\left(\mathit{Frecuencia\ relativa\ acumulada}\right)$	Fra% $\left(\mathit{fra\ porcentual}\right)$
1	112,5	113	217	217,5	165	13	13	0,1083	0,1083	10,83%
2	217,5	218	322	322,5	270	20	33	0,1667	0,275	27,50%
3	322,5	323	427	427,5	375	14	47	0,1167	0,3917	39,17%
4	427,5	428	532	532,5	480	17	64	0,1417	0,5334	53,34%
5	532,5	533	637	637,5	585	16	80	0,1333	0,6667	66,67%
6	637,5	638	742	742,5	690	13	93	0,1083	0,775	77,50%
7	742,5	743	847	847,5	795	16	109	0,1333	0,9083	90,83%
8	847,5	848	952	952,5	900	11	120	0,0917	1	100,00%

UNIDAD TRES
DIAGRAMAS DE FRECUENCIAS

3. INTRODUCCIÓN

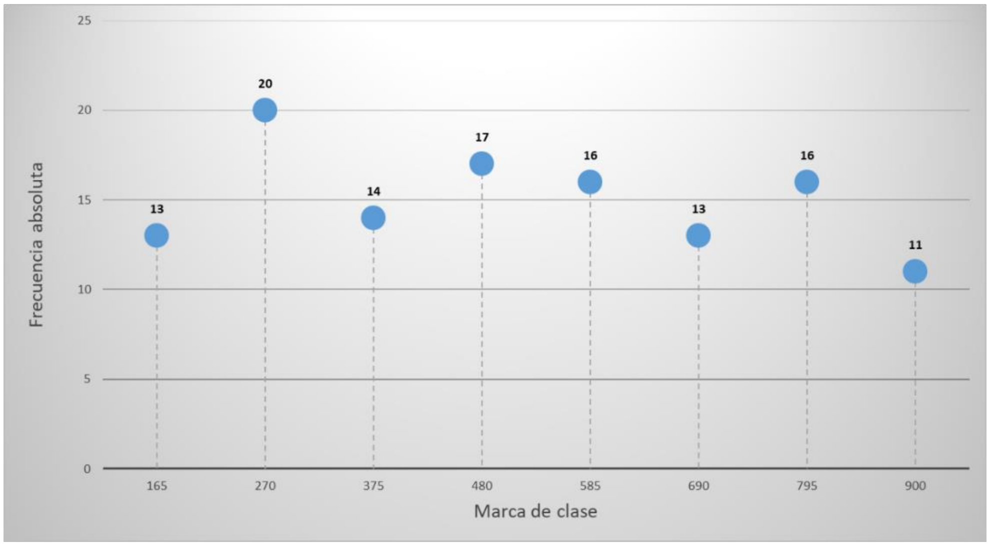
Las gráficas estadísticas son herramientas visuales que permiten representar datos de manera clara y concisa. Estos son útiles para identificar patrones, tendencias y relaciones entre variables, lo que facilita la interpretación y toma de decisiones basadas en datos. La variedad de gráficos existente es muy amplia, y cada uno cuenta con características que nos ayudaran a representar de manera apropiada distintos tipos de variables, no obstante, esta variedad a menudo puede provocar que usemos un tipo de grafico incorrecto provocando que nos alejemos del objetivo esencial de los gráficos estadísticos en cuanto a la consulta y análisis de los datos (Abad, Huapaya, 2009).

Por su importancia, en esta obra centraremos nuestra atención en los diagramas de frecuencias, para lo cual usaremos la distribución de frecuencias por intervalos del caso 2 mostrado en la tabla 23.

3.1 DIAGRAMA DE PUNTOS

El diagrama de puntos es una herramienta gráfica utilizada en estadística para representar la distribución de un conjunto de datos unidimensionales. En este tipo de diagrama, cada punto representa un valor individual en la escala de medición. El eje horizontal del diagrama representa la variable en estudio, mientras que el eje vertical representa la frecuencia absoluta de los valores observados. Es útil para visualizar la forma de la distribución de los datos, identificar valores atípicos y comparar varias distribuciones de datos. Además, es una alternativa simple y rápida a otros tipos de gráficos más complejos como el histograma. Su interpretación depende de la forma y distribución de los puntos, si están agrupados en una zona específica puede indicar una concentración de valores en esa zona, si los puntos están dispersos puede indicar una distribución uniforme o aleatoria de los valores. Si hay algún punto aislado o fuera de la línea principal de puntos puede indicar una observación inusual.

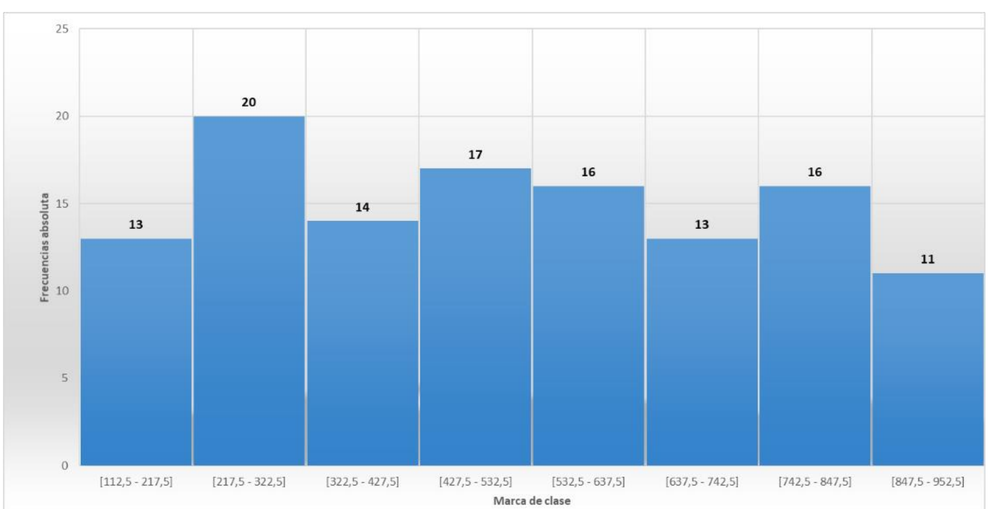
Figura 1. Diagrama de puntos



3.2 HISTOGRAMA DE FRECUENCIA ABSOLUTA

En el histograma, los datos se dividen en intervalos y se cuenta el número de observaciones que caen dentro de cada intervalo. Estas frecuencias se representan en el eje vertical, mientras que los intervalos se colocan en el eje horizontal. Para los histogramas debemos usar los límites reales inferior y superior

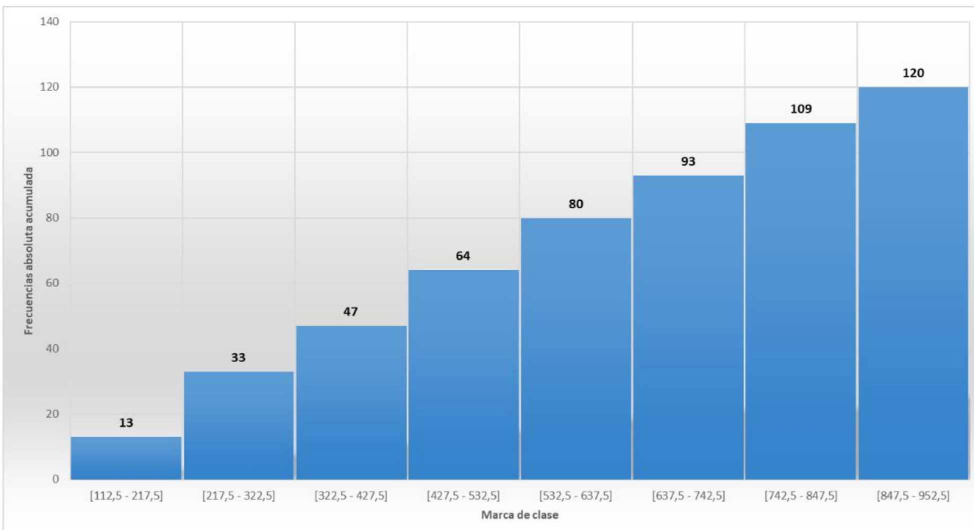
Figura 2. Histograma de frecuencia absoluta



3.3 HISTOGRAMA DE FRECUENCIA ABSOLUTA ACUMULADA

El histograma de frecuencia absoluta acumulada es una herramienta útil para analizar la distribución de datos en un conjunto de observaciones. Este tipo de histograma muestra la frecuencia acumulada de los valores en el eje x, lo que permite visualizar la proporción de valores que se encuentran por debajo o por encima de un cierto valor.

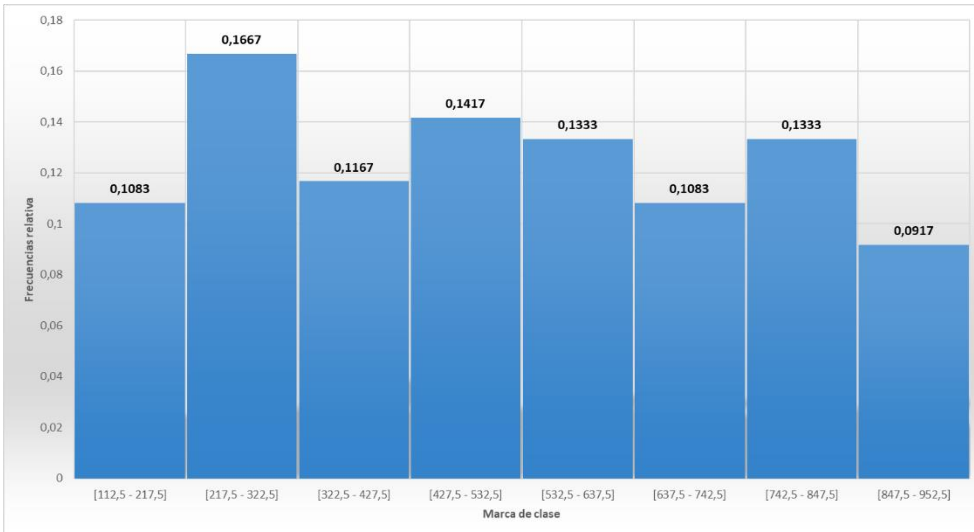
Figura 3. Histograma de frecuencia absoluta acumulada



3.4 HISTOGRAMA DE FRECUENCIA RELATIVA

A diferencia de un histograma de frecuencia absoluta, que muestra el número de veces que ocurre cada valor en un conjunto de datos, un histograma de frecuencia relativa muestra la proporción de veces que ocurre cada valor en un conjunto de datos.

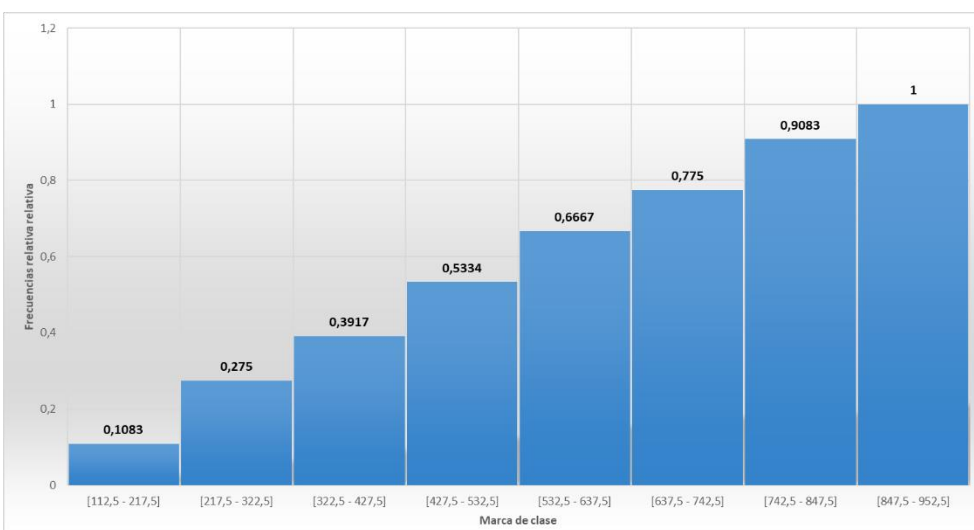
Figura 4. Histograma de frecuencia relativa



3.5 HISTOGRAMA DE FRECUENCIA RELATIVA ACUMULADA

En este tipo de histograma, se muestra la frecuencia acumulada de los datos en lugar de la frecuencia absoluta. Esto permite ver cómo se distribuyen los datos en relación con el total de observaciones.

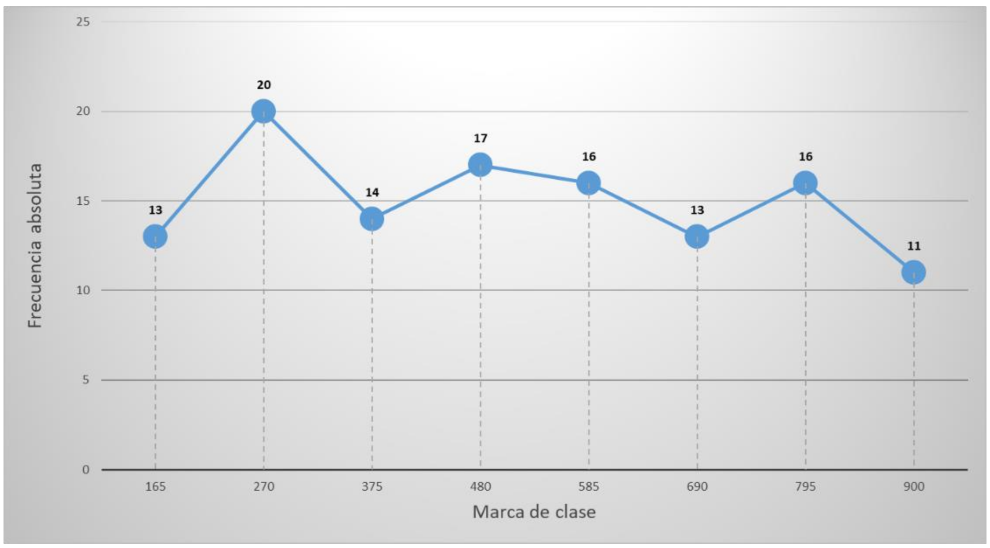
Figura 5. Histograma de frecuencia relativa acumulada



3.6 POLÍGONO DE FRECUENCIA ABSOLUTA

El polígono de frecuencia utiliza segmentos de línea conectados a puntos directamente por encima de los valores del punto medio de la clase. Su grafica es muy similar a un histograma, excepto que usa segmentos de línea en lugar de barras, su forma permite visualizar la distribución de los datos de manera rápida ayudando a identificar valores atípicos o patrones en los datos.

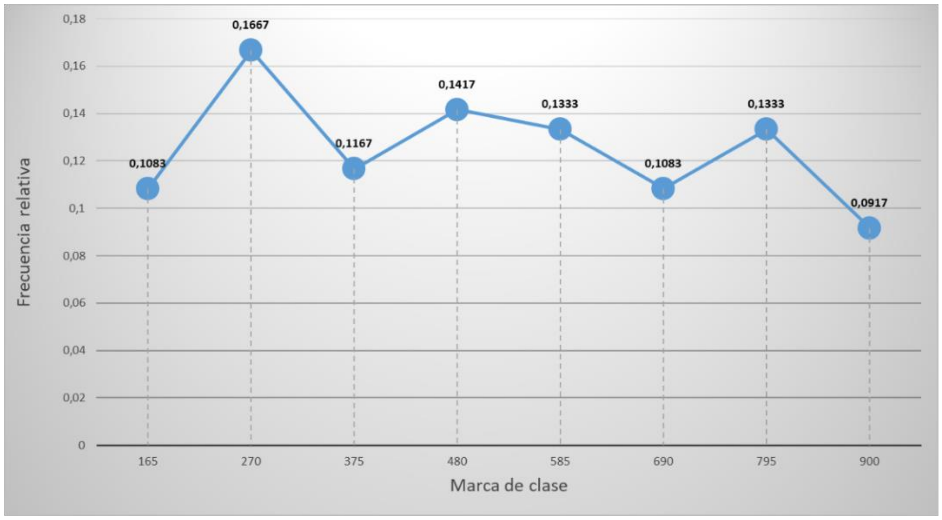
Figura 6. Polígono de frecuencia absoluta



3.7 POLÍGONO DE FRECUENCIA RELATIVA

El polígono de frecuencia relativa es una herramienta útil para la toma de decisiones en procesos empresariales porque permite visualizar la distribución de datos y tomar decisiones basadas en la frecuencia de ocurrencia de ciertos eventos o resultados. Mediante su análisis se puede identificar oportunidades de mejora en los procesos empresariales.

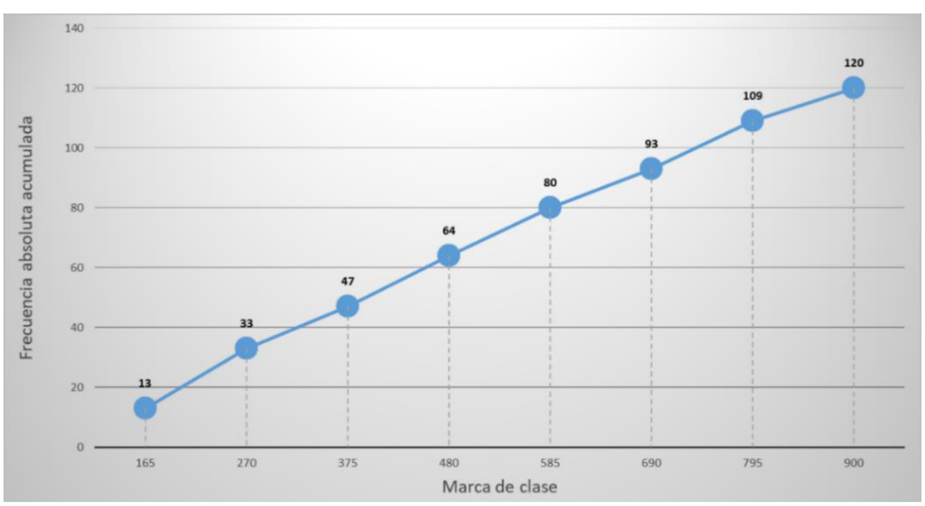
Figura 7. Polígono de frecuencia relativa



3.8 OJIVA DE FRECUENCIA ABSOLUTA

La ojiva de frecuencia absoluta es una herramienta gráfica utilizada en estadística para representar la distribución de frecuencias de una variable acumulada. La igual que los polígonos, esta utiliza segmentos de línea conectados a puntos directamente por encima de los valores del punto medio de la clase. La importancia de la ojiva de frecuencia absoluta radica en su capacidad para proporcionar una visión general de la distribución de los datos, permitiendo identificar patrones y tendencias en los mismos.

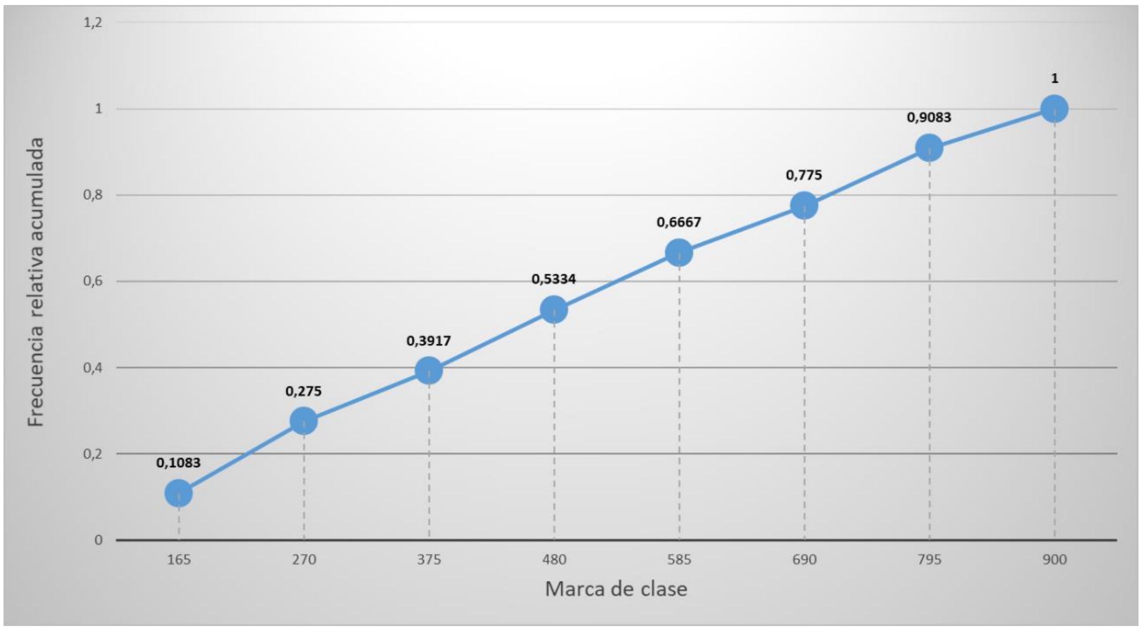
Figura 8. Ojiva de frecuencia absoluta



3.9 OJIVA DE FRECUENCIA RELATIVA

A diferencia de la ojiva de frecuencia absoluta, esta se construye con los valores de frecuencia relativa acumulada.

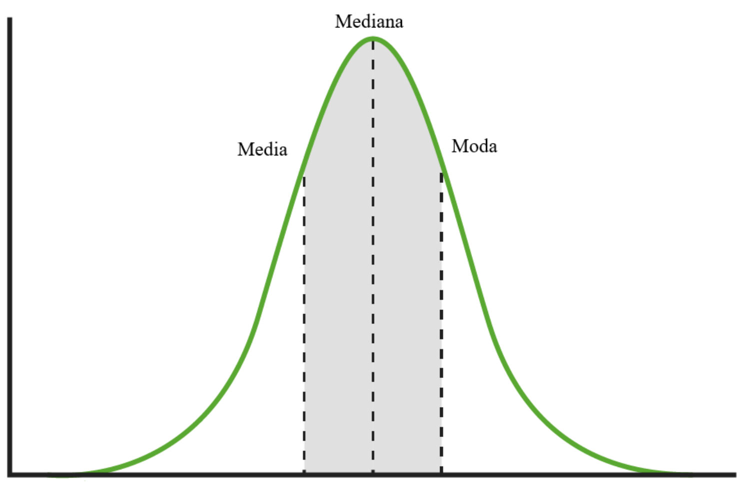
Figura 9. Ojiva de frecuencia relativa



UNIDAD CUATRO
MEDIDAS DE TENDENCIA CENTRAL

4. INTRODUCCIÓN

Las medidas de tendencia central se utilizan en estadística para resumir y describir un conjunto de datos, las tres medidas de tendencia central más comunes son la media, la mediana y la moda, sin embargo, existen otras como la media geométrica y la media armónica, que, aunque no son muy comunes igual siguen siendo útiles. Los valores que toman estas medidas tienden a ubicarse hacia el centro del grupo de datos de la muestra. En una distribución uniforme las medidas de tendencia central en la campana de Gauss se verían de esta forma



4.1 LA MEDIA

La media, es la medida de tendencia central más utilizada y representa el promedio de un conjunto de valores, cuando este conjunto es la población recibe el nombre de *media poblacional* y cuando el conjunto es la muestra recibe el nombre de *media muestra*. Esta medida es ampliamente usada en diversos campos, incluidos las finanzas, la economía y la ciencia. En un modelado predictivo, por ejemplo, la media aritmética se utiliza a menudo para representar el valor esperado de una variable en una población determinada. Se calcula sumando todos los valores de un conjunto de datos y dividiendo por el número de valores.

Su simbología varía dependiendo de la fuente de datos, así, la media poblacional se representa con la letra griega mi en minúscula (μ) y la media muestral con la letra x sombreada ($\bar{x}$). Cabe recalcar, que la fuente de datos no altera el método de cálculo. Las fórmulas de cálculo son las siguientes.

Media aritmética para datos no agrupados

$$\bar{\mathrm{X}} = \frac{\sum x}{N}$$

Media aritmética para datos agrupados

$$\bar{\mathbf{X}} = \frac{\sum(\mathrm{X}_i * \mathrm{fa}_i)}{N}$$

Los resultados siempre se redondeas a la misma cantidad de cifras decimales de la muestra. Para ejemplificar las expresiones usaremos los datos del caso 2, resuelto en el capítulo 2.

Media aritmética de datos no agrupados:

$$\bar{\mathrm{x}} = \frac{113 + 124 + 125 + 135 + \cdots + 950}{120}$$

$$\bar{\mathrm{x}} = \frac{61812}{120}$$

$$\bar{\mathrm{x}} = 515{,}1$$

$$\bar{\mathrm{x}} \cong 515$$

Media aritmética de datos agrupados

$$\bar{\mathrm{x}} = \frac{(165 * 13) + (270 * 20) + (375 * 14) + \cdots + (900 * 11)}{120}$$

$$\bar{\mathrm{x}} = \frac{61905}{120}$$

$\bar{x} = 515{,}875$

$\bar{x} \approx 516$

Como se puede observar en el ejemplo, la media de datos agrupados y no agrupados arrojan valores similares, esto es evidente, considerando que la media de datos no agrupados es el más exacto.

4.2 LA MEDIANA

La mediana es un valor estadístico que se utiliza comúnmente para describir el valor medio de una distribución de datos ordenados. A diferencia de la media, que se ve afectada por valores extremos o atípicos, la mediana es más resistente a estos valores y proporciona una mejor comprensión de la tendencia central de los datos. La mediana no tiene una notación estándar, algunos autores usan símbolos $\tilde{x}$, otros el símbolo $\mu_{1/2}$, por este motivo, en esta obra lo representaremos con las siglas (Me).

El método de cálculo para *datos no agrupados* se calcula por simple inspección, para ello se siguen los siguientes pasos:

Se ordenan los datos de la población o muestra.

Sí, la cantidad de datos es impar, la mediana es igual al dato en la posición N/2 (redondeando). Por ejemplo, en una muestra con 15 datos [1, 1, 3, 3, 5, 6, 6, 9, 11, 11, 12, 12, 13, 14, 17], la mediana es igual al dato en la posición 8 ($15/2 = 7{,}5$), es decir, $\mathrm{Me} = 9$.

Si la cantidad de datos es par, la mediana será igual al promedio de los dos datos centrales, es decir los datos en la posición N/2 y N/2+1, por ejemplo, en una muestra con 16 datos [1, 1, 3, 3, 5, 6, 6, 9, 11, 11, 12, 12, 13, 14, 17, 18] la mediana es

$$Me = \frac{9 + 11}{2} = \frac{20}{2} = 10$$

El método de cálculo para datos agrupados es el siguiente.

a) En la tabla de frecuencias, se identifica la clase que contiene al dato central, es decir la posición N/2.
b) Se aplica la siguiente formula sobre la clase de la mediana.

$$\mathrm{Me} = \mathrm{Xi} + \frac{\frac{\mathrm{N}}{2} - \mathrm{faa}_{-1}}{\mathrm{fa}}\ (\mathrm{Xs} - \mathrm{Xi})$$

Donde Xi es el límite inferior de la clase de la mediana, faa_{-1} es la frecuencia absoluta acumulada de la clase anterior a la mediana, fa es la frecuencia absoluta de la clase de la mediana, y Xs es el límite superior de la clase de la mediana. Para ejemplificar las expresiones usaremos los datos del caso 2, resuelto en el capítulo 2.

Mediana de datos no agrupados

N=120 y este es un número par
Las posiciones
N/2= 60
N/2+1 = 61

Por tanto,

$$Me = \frac{Dato\ en\ la\ posición\ 60 + Dato\ en\ la\ posición\ 61}{2}$$

$$Me = \frac{506 + 506}{2} = 506$$

Mediana de datos agrupados

La posición N/2 = 120/2 =60

Por tanto, la clase que contienen la mediana es la cuarta, esto se puede identificar en la faa como se describe en la tabla.

Núm. de clase	Xi	Xs	faa	
	$\begin{pmatrix} Limite \\ inferior \end{pmatrix}$	$\begin{pmatrix} Limite \\ superior \end{pmatrix}$	$\begin{pmatrix} Frecuencia \\ absoluta \\ acumulada \end{pmatrix}$	
1	113	217	13	Esta clase contienen los datos desde la posición 1 a 13
2	218	322	33	Esta clase contienen los datos desde la posición 14 a 33
3	323	427	47	Esta clase contienen los datos desde la posición 34 a 47
4	428	532	64	Esta clase contienen los datos desde la posición 48 a 64
5	533	637	80	Esta clase contienen los datos desde la posición 65 a 80
6	638	742	93	Esta clase contienen los datos desde la posición 81 a 93
7	743	847	109	Esta clase contienen los datos desde la posición 94 a 109
8	848	952	120	Esta clase contienen los datos desde la posición 110 a 120

Finalmente, aplicamos la fórmula de la mediana.

Núm. clase	Xri	Xi	Xs	Xrs	X	fa	faa	fr	
: :	: :	: :	: :	: :	: :	: :	: :	: :	
3	322,5	323	427	427,5	375	14	47	0,1167	...
4	427,5	428	532	532,5	480	17	64	0,1417	...
: :	: :	: :	: :	: :	: :	: :	: :	: :	

$$\text{Me} = 428 + \frac{\frac{120}{2} - 47}{17}\,(532 - 428)$$

$$\text{Me} = 507,52$$

$$\text{Me} \approx 508$$

Observe que la mediana de datos agrupados y no agrupados arrojan valores similares. Recuerde que la mediana de datos no agrupados es el valor mas más exacto.

4.3 LA MODA

La moda es una medida de tendencia central utilizada en estadística para describir el valor más común o frecuente en un conjunto de datos. A diferencia de la media y la mediana, la moda no se ve afectada por valores extremos o atípicos en los datos. La forma más común de representar la moda es a través de las siglas Mo.

El método de cálculo para *datos no agrupados* se lo realiza por simple inspección, observando el dato con mayor frecuencia, es decir, cuál de los datos se repite mayor cantidad de veces. Por ejemplo, en una muestra con 10 datos [1, 1, 2, 2, 2, 5, 6, 6, 9, 13] la Mo=2 porque este dato se repite tres veces y no existe otro dato que repita mayor cantidad. En el caso de existir dos números con mayor frecuencia se dice que el ejercicio es bimodal y se apunta los dos valores como moda 1 (Mo1) y moda 2 (Mo2), de existir tres números con mayor frecuencia se habla de un ejercicio trimodal y se apuntas los tres valores como modas, etc.

El método de cálculo para *datos agrupados* sigue el siguiente procedimiento:

a) Se asume que la moda se encuentra en la clase con mayor frecuencia absoluta, justamente por ser el grupo con mayor cantidad de datos. A esta se la llama “clase de la moda”.
b) Posterior se aplica la siguiente regla.

$$Mo = Xi + \frac{\Delta_1}{\Delta_1 + \Delta_2} * AIC$$

donde

$\Delta_1 = fa - fa_{-1}$ (fa_{-1} es la fa de la clase anterior a la moda)

$\Delta_2 = fa - fa_{+1}$ (fa_{+1} es la fa de la clase posterior a la moda)

Resolviendo la ecuación tenemos

$$Mo = Xi + \frac{fa - fa_{-1}}{2fa - fa_{-1} - fa_{+1}} * (Xs - Xi)$$

En el caso en que la clase de la moda sea la primera de la tabla, la fa_{-1} se apunta con el valor de cero, y lo mismo si la clase de la moda es la última de la tabla la fa_{+1} también es igual a cero. Para ejemplificar las expresiones usaremos los datos del caso 2, resuelto en el capítulo 2.

Moda de datos no agrupados

Observando la muestra se puede determinar que, los valores con mayor frecuencia son 311 y 854 repitiéndose tres veces cada uno como se muestra a continuación.

113	197	239	311	384	456	506	557	659	724	806	854
124	197	256	318	390	456	517	567	665	729	809	854
125	198	258	319	390	470	519	581	668	730	817	854
135	219	261	324	401	480	523	588	677	746	822	865
156	220	283	327	406	480	533	591	682	760	824	880
157	221	304	331	411	481	543	596	686	774	833	907
158	224	308	346	418	484	545	614	698	782	837	911
163	225	310	358	442	485	545	621	709	787	845	918
172	227	311	359	445	500	549	628	713	798	845	927
186	238	311	369	447	506	549	634	724	799	848	950

Por tanto.

Mo1 = 311

Mo2 = 854

Moda de datos agrupados

Se observa que la clase con mayor frecuencia absoluta es la segunda, como se muestra a continuación:

Núm de clase	Xri	Xi	Xs	Xrs	X	fa	
1	112,5	113	217	217,5	165	13	…
2	217,5	218	322	322,5	270	20	…
3	322,5	323	427	427,5	375	14	…
4	427,5	428	532	532,5	480	17	…
5	532,5	533	637	637,5	585	16	…
6	637,5	638	742	742,5	690	13	…
7	742,5	743	847	847,5	795	16	…
8	847,5	848	952	952,5	900	11	…

Clase de la moda

Aplicando la fórmula a la clase de la moda, quedaría de la siguiente forma.

Núm. de clase	Xi	Xs	X	fa	faa	
1	113	217	165	13	13	…
2	218	322	270	20	33	…
3	323	427	375	14	47	…
: :	: :	: :	: :	: :	: :	

$$Mo = 218 + \frac{20 - 13}{2 * 20 - 13 - 14} * (322 - 218)$$

$$Mo = 274$$

Observando los resultados podemos concluir que las modas de datos no agrupados y agrupados raramente van a coincidir, esto se debe a que en la tabla de frecuencias no se tienen los datos como tal a fin de observar cual es el valor que más se repite, por tanto, su estimación es supuesta.

4.4 LA MEDIA GEOMÉTRICA

Media Geométrica es una fórmula matemática que se utiliza para calcular el promedio de un conjunto de números multiplicados en raíz enésima. Esta fórmula se ha utilizado en diversos campos a lo largo de la historia, por ejemplo, en finanzas se utiliza para calcular el retorno promedio de la inversión durante un período de tiempo, esto es útil para los inversores que desean conocer el rendimiento general de su cartera teniendo en cuenta el efecto compuesto de los rendimientos a lo largo del tiempo, en demografía se utiliza para calcular la tasa de crecimiento promedio de una población durante un período de tiempo, esto es útil para los formuladores que desean comprender la tendencia general del crecimiento demográfico y planificar consecuencias futuras. La forma más común de representar la media geométrica es con la letra G mayúscula.

Las fórmulas de cálculo son las siguientes:

Media geométrica para datos no agrupados

$$G = \sqrt[N]{X_1 * X_2 * X_3 * X_4 * \ldots . * X_N}$$

Cuando se usa calculadora de mano, por lo general estas no permiten ingresar directamente todos los valores de la muestra, en estos casos se aplica la propiedad producto de raíces enésimas, esto es.

$$G = \sqrt[N]{X_1 * X_2 * \dots * X_n} \;*\; \sqrt[N]{X_{n+1} * X_{n+2} * \dots * \; X_m} \;*\; \sqrt[N]{X_{m+1} * X_{m+2} * \dots * X_N}$$

Media geométrica para datos agrupados

$$G = \sqrt[N]{{X_1}^{f_1} * {X_2}^{f_2} * {X_3}^{f_3} * \; {X_4}^{f_4} * \dots \; * {X_N}^{f_N}}$$

Los resultados siempre se redondeas a la misma cantidad de cifras decimales de la muestra. Para ejemplificar las expresiones usaremos los datos del caso 2, resuelto en el capítulo 2.

Media geométrica de datos no agrupados

$$G = \sqrt[120]{113 * 124 * 125 * 135 * 156 * \dots * 950}$$

$$G = 452,478$$

$$G \approx 553$$

Media geométrica de *datos agrupados*

$$G = \sqrt[120]{165^{13} * 270^{20} * 375^{14} * \; 480^{17} * \dots \; * 900^{11}}$$

$$G = 456,686$$

$$G \approx 457$$

Al igual que la media y la mediana, la media geométrica de datos no agrupados y agrupados deben ser similares.

4.5 LA MEDIA ARMÓNICA

La media armónica es una medida de tendencia central que se utiliza en estadística para calcular el promedio de un conjunto de datos. A diferencia de la media aritmética y la mediana, la media armónica se utiliza principalmente para datos que varían ampliamente, su precisión está limitado en por dos aspectos.

El valor de la media armónica puede ser susceptible a valores extremos o atípicos en un conjunto de datos, lo que puede afectar su precisión y relevancia en el análisis estadístico.

La media armónica también puede ser influenciada por la presencia de ceros o valores negativos en el conjunto de datos, lo que puede afectar su capacidad para proporcionar una representación precisa de la media del conjunto de datos.

Su interpretación también conlleva dos consideraciones importantes a tomar en cuenta.

El uso exclusivo de la media armónica puede llevar a una subestimación de la media real del conjunto de datos, lo que puede sesgar los resultados del análisis estadístico.

La media armónica también puede sesgar los resultados del análisis estadístico si se utiliza como la única medida de tendencia central, ya que no proporciona información sobre la distribución de los datos.

La forma más común de representar la media geométrica es con la letra H mayúscula. Las fórmulas de cálculo son las siguientes:

Media armónica para datos no agrupados

$$\mathrm{H} = \frac{\mathrm{N}}{\sum\left(\frac{1}{X_i}\right)}$$

Media armónica para datos agrupados

$$H = \frac{N}{\sum\left(\frac{fa_i}{X_i}\right)}$$

Los resultados siempre se redondeas a la misma cantidad de cifras decimales de la muestra. Para ejemplificar las expresiones usaremos los datos del caso 2, resuelto en el capítulo 2.

Media geométrica de *datos no agrupados*

$$H = \frac{120}{\frac{1}{113} + \frac{1}{124} + \frac{1}{125} + \frac{1}{135} + \cdots + \frac{1}{950}}$$

$$H = 384,481$$

$$H \approx 385$$

Media geométrica de *datos agrupados*

$$H = \frac{120}{\frac{13}{165} + \frac{20}{270} + \frac{14}{375} + \cdots + \frac{11}{900}}$$

$$H = 394{,}54$$

$$H \approx 395$$

Al igual que la media, la mediana y la media geométrica, la media armónica de datos no agrupados y agrupados deben ser similares.

UNIDAD CINCO
TEORÍA ELEMENTAL DE PROBABILIDADES

5. INTRODUCCIÓN.

La teoría de la probabilidad estadística es una rama de las matemáticas que se ocupa del estudio de eventos aleatorios y su probabilidad de ocurrencia.

El uso de la probabilidad para medir la incertidumbre y la variabilidad se remonta a cientos de años, la probabilidad ha encontrado aplicaciones en áreas tan diversas como la medicina, los juegos de azar, la predicción meteorológica y otras; el concepto de azar e incertidumbre es tan antiguo como la civilización misma, los registros históricos muestran como los humanos han luchado contra la incertidumbre del clima, su fascinación por los juegos de azar, etc. (DeGroot, Schervish).

A continuación, algunos ejemplos sobre la aplicación de la teoría de la probabilidad en la actualidad.

Finanzas: Evaluar el riesgo y el rendimiento de la inversión, modelar los precios de las acciones y analizar datos financieros.

Física: Modelar el comportamiento de partículas, analizar la mecánica cuántica y predecir los resultados de los experimentos.

Ingeniería: Diseñar experimentos, probar materiales y analizar datos de experimentos.

Ciencias sociales: Analizar datos de encuestas, modelar el comportamiento humano y predecir los resultados de los fenómenos sociales.

5.1 DEFINICIÓN CLÁSICA DE PROBABILIDAD

Según Chalmers (2018), dado un evento E con n posibles formas de ocurrir, y de todas estas solo algunas (h) cumplen una cierta condición buscada (a esto llamaremos éxito), la probabilidad de que se suscite el evento E se calcula de la siguiente manera:

$$Pr\{E\} = p = \frac{h}{n}$$

Y la probabilidad de que no ocurra el evento (a esto llamaremos fracaso) se denota como Pr{no E} o simplemente q, cuyo valor se calcula de la siguiente manera:

$$Pr\{no\ E\} = q = \frac{n-h}{n}$$

De lo anterior también se puede deducir las siguientes relaciones entre p y q

$$q = 1 - \frac{h}{n}$$

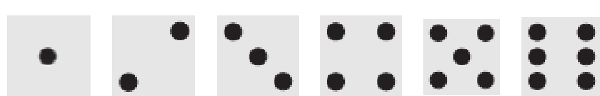
$$q = 1 - p$$

EJEMPLO 1: Al lanzar un dado ¿Cuál es la probabilidad de que salga un numero par?

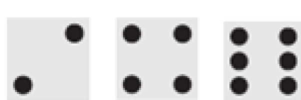
SOLUCIÓN

Datos:

- E = lanzar un dado
- n= 6 (son las formas posibles que tiene el dado)



- h= 3 (de las 6 posibles formas, solo 3 cumplen la condición de ser par)



Solución:

$$p = \frac{3}{6}$$

$$p = \frac{1}{2} = 0{,}5$$

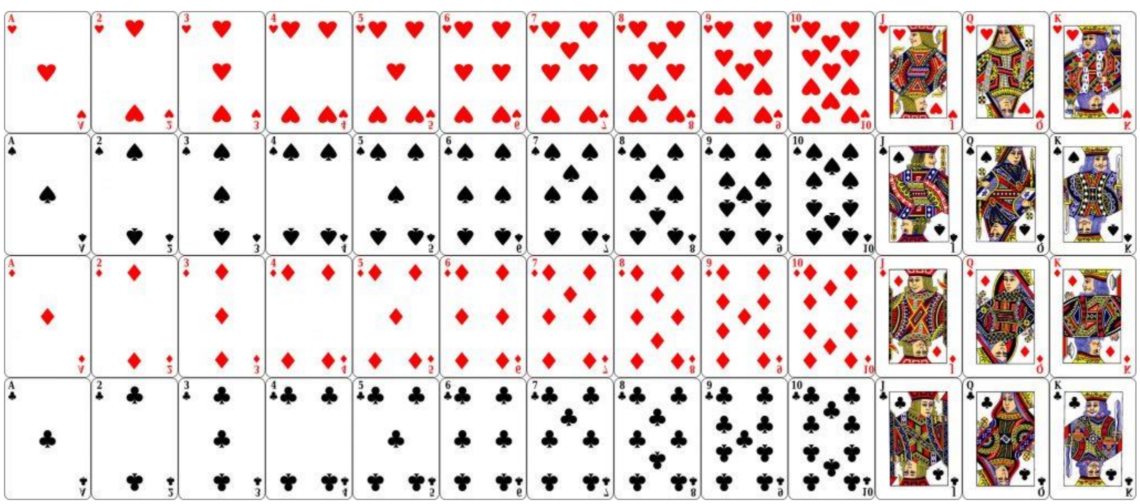
EJEMPLO 2: De una baraja de 52 naipes se extrae una carta al azar. Cuál es la probabilidad de:

a) La carta extraída sea un trébol.
b) La carta extraída no sea un As.
c) La carta extraída no sea un trébol.

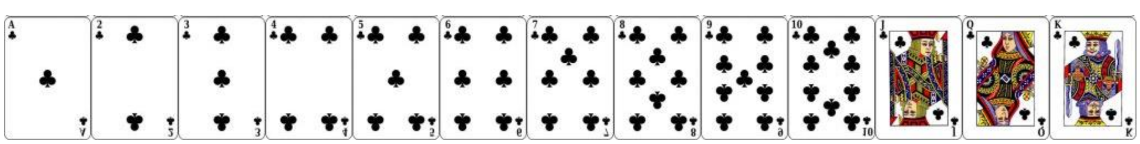
SOLUCIÓN LITERAL A:

Datos:

- E = Extraer una carta al azar
- n= 52 (son todas las formas posibles que tiene la baraja)



- h= 13 (de las 52 cartas 13 son tréboles)



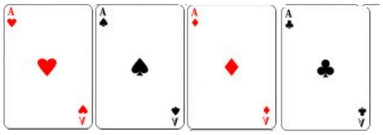
Solución:

$$p = \frac{13}{52}$$

$$p = \frac{1}{4} = 0,25$$

SOLUCIÓN LITERAL B:

Considerando que el naipe de 52 cartas contiene cuatro As



La solución es la siguiente:

$$Pr\{no\ E\} = q = \frac{52-4}{52} = \frac{12}{13} = 0{,}92$$

SOLUCIÓN LITERAL C:

Si sabemos que la probabilidad de que al extraer una carta de trébol es 0,25. Entonces la probabilidad de que no sea trébol se puede calcular de la siguiente manera.

$$q = 1 - 0{,}25 = 0{,}75$$

5.2 PROBABILIDAD CONDICIONAL: EVENTOS INDEPENDIENTES Y DEPENDIENTES

Según Hogg, McKean y Craig (2019), si E1 y E2 son dos eventos, la probabilidad de que ocurra E2, dado que E1 ha ocurrido, se denota Pr{E2|E1} o Pr{E2 dado E1} y se conoce como la probabilidad condicional de E2 dado que E1 ha ocurrido. En función a lo descrito se puede distinguir dos casos:

- Si la ocurrencia del E1 no afecta la probabilidad de ocurrencia del E2, entonces se dice que E1 y E2 son eventos independientes
- Por el contrario, si la ocurrencia del E1 si afecta la probabilidad de ocurrencia del E2, entonces se dice que E1 y E2 son eventos dependientes.

Si se denota con E_1E_2 el evento de que "tanto E1 como E2 ocurran", evento al que suele llamarse evento compuesto, entonces.

Pr{E1E2} = Pr{E1} * Pr{E2]   Para sucesos independientes

Pr{E1E2} = Pr{E1} * Pr{E2|E_1}   Para sucesos dependientes

Las fórmulas anteriores se pueden generalizar para n eventos, de la siguiente manera.

$Pr\{E1\ E2 \ldots E_n\} = Pr\{E1\} * Pr\{E2\} * \ldots * Pr\{En\}$ Para sucesos independientes

$Pr\{E1\ E2 \ldots En\} = Pr\{E1\} * Pr\{E2|E_1\} * \ldots * Pr\{En|E_1 E_{2 \ldots} E_{n-1}\}$ Para sucesos dependientes.

EJEMPLO 3: Hallar la probabilidad de que al extraer dos cartas de una naipe, el primero sea un As y el segundo sea un corazón rojo. Hay que considerar que después de extraer la primera carta, esta se devuelve al naipe antes de extraer la segunda carta.

SOLUCIÓN:

Como el ejercicio especifica que se devuelve la carta, entonces se habla de sucesos independientes. Entonces:

DATOS:

- E_1: Extraer la primera carta
- E_2: Extraer la segunda carta (después de reponer la primera)
- n=52 cartas
- h_1: 4 (hay 4 As en el naipe)
- h_2: 13 (hay 13 corazones rojos en el naipe)

Entonces:

$$Pr\{E1E2\} = Pr\{E1\} * Pr\{E2\}$$

$$Pr\{E1E2\} = \frac{h_1}{n} * \frac{h_2}{n}$$

$$Pr\{E1E2\} = \frac{4}{52} * \frac{13}{52}$$

$$Pr\{E1E2\} = 0{,}0192$$

EJEMPLO 4: Considerar el mismo EJEMPLO 3 tomando en cuenta que después de extraer la primera carta, esta **NO** se devuelve al naipe.

SOLUCIÓN:

Como el ejercicio especifica que **NO** se devuelve la carta, entonces se habla de sucesos independientes. Entonces:

DATOS:

- E_1: Extraer la primera carta
- E_2: Extraer la segunda carta (No se repuso la primera carta a la baraja)
- n_1=52 cartas
- n_2=51 cartas
- h_1: 4 (hay 4 As en el naipe)
- h_2: 13 (hay 13 corazones rojos en el naipe)

Entonces:

$$Pr\{E1E2\} = Pr\{E1\} * Pr\{E2|E_1\}$$

$$Pr\{E1E2\} = \frac{h_1}{n_1} * \frac{h_2}{n_2}$$

$$Pr\{E1E2\} = \frac{4}{52} * \frac{13}{51}$$

$$Pr\{E1E2\} = 0{,}0196$$

EJEMPLO 5: Supóngase que una urna de sorteos contiene 3 bolas rojas y 2 bolas azules. Sea $E1$ el evento "la primera bola que se saca es azul" y $E2$ el evento "la segunda bola que se saca es azul", donde las bolas no se vuelvan a colocar en la urna una vez sacadas.

SOLUCIÓN

Aquí $E1$ y $E2$ son eventos dependientes.

DATOS:

- E_1: La primera bola que se saca es azul.
- E_2: La segunda bola que se saca es azul.
- n_1=5 bolas en total
- n_2=4 bolas en total
- h_1: 2 (hay 2 bolas azules)
- h_2: 1 (Suponiendo que el primer evento es éxito, entonces solo quedaría una bola azul en la urna)

Entonces:

$$Pr\{E1E2\} = Pr\{E1\} * Pr\{E2|E_1\}$$

$$Pr\{E1E2\} = \frac{h_1}{n_1} * \frac{h_2}{n_2}$$

$$Pr\{E1E2\} = \frac{2}{5} * \frac{1}{4}$$

$$Pr\{E1E2\} = \frac{1}{10} = 0{,}1$$

5.3 EVENTOS MUTUAMENTE EXCLUYENTES

Los eventos mutuamente excluyentes son aquellos en los que puede ocurrir un evento E_1 o un evento E_2 pero no los dos al mismo tiempo. Su cálculo es el siguiente:

$$Pr\{E_1+E_2\} = Pr\{E_1\} + Pr\{E_2\}$$

Por el contrario, si en un determinado suceso se pudiera obtener evento E1, E2, o los dos al mismo tiempo, entonces su cálculo es el siguiente:

$$Pr\{E_1+E_2\} = Pr\{E_1\} + Pr\{E_2\} - Pr\{ E_1E_2\}$$

EJEMPLO 6: Si E_1 es el evento "de una baraja se extrae un as" y E_2 es el evento "de una baraja se extrae un rey", Calcular la probabilidad de que al extraer una carta al azar, esta sea un As o un Rey.

SOLUCIÓN

En este caso al extraer una carta esta puede ser un As o un Rey pero no ambas a la vez, porque de las 52 cartas ninguna es As y Rey al mismo tiempo, por tanto cualquiera que fuera la carta que saquemos nunca se va a cumplir las dos condiciones al mismo tiempo.

A♣	2♣	3♣	4♣	5♣	6♣	7♣	8♣	9♣	10♣	J♣	Q♣	K♣
A♦	2♦	3♦	4♦	5♦	6♦	7♦	8♦	9♦	10♦	J♦	Q♦	K♦
A♥	2♥	3♥	4♥	5♥	6♥	7♥	8♥	9♥	10♥	J♥	Q♥	K♥
A♠	2♠	3♠	4♠	5♠	6♠	7♠	8♠	9♠	10♠	J♠	Q♠	K♠

Por tanto, E_1 y E_2 son eventos mutua mente excluyentes, y su cálculo es el siguiente.

$$Pr\{E_1+E_2\} = Pr\{E_1\} + Pr\{E_2\}$$

$$Pr\{E_1+E_2\} = \frac{4}{52} + \frac{4}{52}$$

$$Pr\{E_1+E_2\} = \frac{2}{13} = 0{,}1538$$

EJEMPLO 7: Si E_1 es el evento "de una baraja se extrae un as" y E_2 es el evento "extraer una espada", Calcular la probabilidad de que, al extraer una carta al azar, esta sea un As o una espada.

SOLUCIÓN

En este caso al extraer una carta esta puede ser un As o una espada, pero al contrario del ejemplo anterior si existe una carta que cumpla las dos condiciones al mismo tiempo, es decir que sea un AS de espadas.

A♣	2♣	3♣	4♣	5♣	6♣	7♣	8♣	9♣	10♣	J♣	Q♣	K♣
A♦	2♦	3♦	4♦	5♦	6♦	7♦	8♦	9♦	10♦	J♦	Q♦	K♦
A♥	2♥	3♥	4♥	5♥	6♥	7♥	8♥	9♥	10♥	J♥	Q♥	K♥
A♠	2♠	3♠	4♠	5♠	6♠	7♠	8♠	9♠	10♠	J♠	Q♠	K♠

Por tanto, E_1 y E_2 **NO** son eventos mutua mente excluyentes, y su cálculo es el siguiente.

$$Pr\{E_1+E_2\} = Pr\{E_1\} + Pr\{E_2\} - Pr\{E_1E_2\}$$

$$Pr\{E_1+E_2\} = \frac{4}{52} + \frac{13}{52} - \frac{1}{52}$$

$$Pr\{E_1+E_2\} = \frac{2}{13} = 0{,}1346$$

6. REFERENCIAS

- Pérez, R. (2010). *Nociones básicas de estadística*. Oviedo, España: Universidad de Oviedo.
- Agresti, A., Finlay, B., De Battisti, F., y Porro, F. (2009). *Statistica*. Milán, Italia: Pearson Italia.
- Triola, M. (2009). *Estadística* (10va Ed.). Ciudad de México, México: Pearson.
- Triola, M. (2018). *Estadística* (12va Ed.). Ciudad de México, México: Pearson.
- Fernández, S., Cordoba, A., y Cordero, J. (2002). Estadística descriptiva. Madrid, España: ESIC editorial.
- Dousbebés, A. (2021). *Estadística aplicada a psicología y educación.* Quito, Ecuador: Centro de publicaciones de la Pontifica Universidad Católica del Ecuador.
- Martínez, C. (2020). *Estadística básica aplicada* (5ta Ed.). Bogotá, Colombia: ECOE Ediciones.
- Posada, G. (2016). *Elementos básicos de estadística descriptiva para el análisis de datos*. Medellín, Colombia: Editorial Luis Amigó.
- Ramachandran, K., y Tsokos. (2009). *Mathematical Statistics with Applications*. San Diego, EEUU: Elsevier Academic Press.
- De Souza, N., Pinheiro, A., y Bastos, B. (2018). *Estatística Inferencial*. Belo Horizonte, Brasil: Editorial Poisson.
- Devore, J. (2008). *Probabilidad y Estadística para Ingeniería y Ciencias*. Ciudad de Mexico, Mexico: Cengage Learning.
- Luzardo, M., Jiménez, M. (2018). *Manual de inferencia estadística*. Medellín, Colombia: Editorial Universidad Pontificia Bolivariana.
- Abad, P, y Huapaya, E. (2009). Guía para la presentación de gráficos estadísticos. Lima, Perú: Instituto Nacional de Estadística e Informática.
- Chalmers, D. (2018). Probability & Statistics 1, Coursebook. Cambridge, United Kingdom: Cambridge University Press
- DeGroot, M., Schervish, M. (2012). Probability and Statistics (4ta. Ed.). United States of America: Pearson Education.
- Hogg, R., McKean, J., y Craig. (2019). Introduction to Mathematical Statistics. United States of America: Pearson Education.

CONCLUSIONES

El libro "Estadística y Probabilidades" presenta una contribución significativa en el campo de la estadística aplicada al introducir el "Método Freire" para la creación de tablas de frecuencia por intervalos. Este método fue probado por un lapso de siete años, y fue diseñado para ser fácilmente implementado mediante herramientas como Microsoft Excel, promueve un enfoque más sistemático, preciso y accesible para el análisis de datos en diversos contextos. Además, la incorporación de conceptos fundamentales de probabilidad enriquece el texto al proporcionar una base sólida para la comprensión y aplicación de análisis estadísticos más complejos. A lo largo de este libro, se demuestra cómo la estadística sirve como una herramienta esencial para interpretar fenómenos y tomar decisiones informadas en campos tan variados como la medicina, la economía, la educación y la tecnología. Al combinar teoría y práctica, el texto empodera a los lectores para abordar problemas reales con rigor analítico, destacando la importancia de adherirse a principios éticos en el manejo y presentación de datos.

RECOMENDACIONES

Se recomienda que el lector utilice este libro como una guía práctica para aplicar técnicas estadísticas en contextos profesionales y académicos. En particular, el "Método Freire" se presenta como una herramienta innovadora que puede ser integrada en proyectos de investigación, análisis de mercado y estudios académicos. Se sugiere implementar este método en ejercicios prácticos utilizando datos reales para familiarizarse con su aplicación y versatilidad. Asimismo, se aconseja que los educadores incorporen los capítulos del libro como material de referencia en cursos de estadística básica y avanzada. Esto puede enriquecer el aprendizaje al conectar los principios estadísticos con aplicaciones concretas que los estudiantes enfrentarán en sus respectivas disciplinas.

La esencia del progreso de un hombre no se mide únicamente por sus logros, sino por el viaje interior que conlleva superar los desafíos del camino. Cada paso dado nos llena de sabiduría y nos prepara para alcanzar nuevas metas, incluso cuando vientos de adversidad soplen en nuestra contra.

Luis Freire Sánchez.

ISTSGA
SAN GABRIEL
INSTITUTO SUPERIOR TECNOLÓGICO

EDICA XXI
Editorial Académica Internacional