



## ANTOLOGIA

**MAESTRIA EN ADMINISTRACION DE LOS**

**SISTEMAS DE SALUD MATERIA:**

**TENDENCIAS Y SISTEMAS DE SALUD EN**  
**MEXICO**

ENERO 2021

**Mtra. Ana Cecilia Gutiérrez Castellanos.**

---

## Marco Estratégico de Referencia

---

### Antecedentes históricos

Nuestra Universidad tiene sus antecedentes de formación en el año de 1978 con el inicio de actividades de la normal de educadoras —Edgar Robledo Santiago , que en su momento marcó un nuevo rumbo para la educación de Comitán y del estado de Chiapas. Nuestra escuela fue fundada por el Profesor Manuel Albores Salazar con la idea de traer educación a Comitán, ya que esto representaba una forma de apoyar a muchas familias de la región para que siguieran estudiando.

En el año 1984 inicia actividades el CBTiS Moctezuma Ilhuicamina, que fue el primer bachillerato tecnológico particular del estado de Chiapas, manteniendo con esto la visión en grande de traer educación a nuestro municipio, esta institución fue creada para que la gente que trabajaba por la mañana tuviera la opción de estudiar por la tarde.

La Maestra Martha Ruth Alcázar Mellanes es la madre de los tres integrantes de la familia Albores Alcázar que se fueron integrando poco a poco a la escuela formada por su padre, el Profesor Manuel Albores Salazar; Víctor Manuel Albores Alcázar en julio de 1996 como chofer de transporte escolar, Karla Fabiola Albores Alcázar se integró en la docencia en 1998, Martha Patricia Albores Alcázar en el departamento de cobranza en 1999.

En el año 2002, Víctor Manuel Albores Alcázar formó el Grupo Educativo Albores Alcázar S.C. para darle un nuevo rumbo y sentido empresarial al negocio familiar y en el año 2004 funda la Universidad Del Sureste.

La formación de nuestra Universidad se da principalmente porque en Comitán y en toda la región no existía una verdadera oferta educativa, por lo que se veía urgente la creación de una institución de educación superior, pero que estuviera a la altura de las exigencias de los jóvenes que tenían intención de seguir estudiando o de los profesionistas para seguir preparándose a través de estudios de posgrado.

Nuestra universidad inició sus actividades el 19 de agosto del 2004 en las instalaciones de la 4ª avenida oriente sur no. 24, con la licenciatura en puericultura, contando con dos grupos de cuarenta alumnos cada uno. En el año 2005 nos trasladamos a las instalaciones de carretera Comitán – Tzimol km. 57 donde actualmente se encuentra el campus Comitán y el corporativo UDS, este último, es el encargado de estandarizar y controlar todos los procesos operativos y educativos de los diferentes campus, así como de crear los diferentes planes estratégicos de expansión de la marca.

## **Misión**

Satisfacer la necesidad de educación que promueva el espíritu emprendedor, basados en Altos Estándares de calidad Académica, que propicie el desarrollo de estudiantes, profesores, colaboradores y la sociedad.

## Visión

Ser la mejor Universidad en cada región de influencia, generando crecimiento sostenible y ofertas académicas innovadoras con pertinencia para la sociedad.

## Valores

- Disciplina
- Honestidad
- Equidad
- Lealtad
- Responsabilidad
- Respeto

## Escudo



El escudo del Grupo Educativo Albores Alcázar S.C. está constituido por tres líneas curvas que nacen de izquierda a derecha formando los escalones al éxito. En la parte superior está situado un cuadro motivado de la abstracción de la forma de un libro abierto.

### Eslogan

—Pasión por Educar

### Balam



Es nuestra mascota, su nombre proviene de la lengua maya cuyo significado es jaguar. Su piel es negra y se distingue por ser líder, trabaja en equipo y obtiene lo que desea. El ímpetu, extremo valor y fortaleza son los rasgos que distinguen a los integrantes de la comunidad UDS.

---

## **TENDENCIAS Y SISTEMAS DE SALUD EN MEXICO**

---

### **OBJETIVO DE LA MATERIA.**

Al finalizar el curso, el estudiante será capaz de aplicar los diversos tópicos matemáticos en el planteamiento y resolución de problemas del ámbito empresarial, a través del uso e interpretación de modelos que le permitirán sustentar y apoyar el proceso de la toma de decisiones.

### **TEMAS**

#### **SEMANA UNO**

#### **ENCUADRE ( ORGANIZACIÓN DE LA MATERIA Y NORMATIVA SOBRE EL PROCESO DE EVALUACION)**

#### **UNIDAD I ESTADÍSTICA DESCRIPTIVA**

**1.1.- Procesamiento estadístico de datos (recolección, organización, presentación, análisis e interpretación de datos).**

**1.2.- Distribuciones de frecuencias.**

**1.3.- Presentación gráfica.**

**1.4.- Medidas de tendencia central.**

**1.5.- Medidas de dispersión.**

**1.6.- Teorema de Tchebyshev.**

**1.7.- Regla empírica.**

## **UNIDAD II TEORÍA DE LA PROBABILIDAD**

### **2.1.- Introducción.**

#### **2.1.1.- Enfoques de probabilidad.**

#### **2.1.2.- Espacio muestral.**

#### **2.1.3.- Eventos simples y compuestos.**

#### **2.1.4.- Leyes de probabilidad.**

#### **2.1.5.- Tablas de contingencia.**

#### **2.1.6.- Teorema de Bayes.**

## **BIBLIOGRAFIA SUGERIDA**

- Enrique y Guijarro, Evolución y Reforma del sistema de salud en México, Editorial Cepal, 2000
- GIEDION, Manuela Villar, Ávila Adriana, Los sistema de salud en Latinoamérica y el papel del seguro Privado, Editorial Fundación Mapre, 2010.
- SANCHES, León Gregorio, Editorial Cárdenas, 2006.

## **CRITERIOS, PROCEDIMIENTOS DE EVALUACION Y ACREDITACION**

**Trabajos 50%**

**Foros 30%**

**Exámen 20% Total 100%**

**Escala de calificación 08- 10**

**Mínima aprobatoria 8**

## **SEMANA 1**

### **UNIDAD I**

#### **ESTADISTICA DESCRIPTIVA**

##### **1.1 PROCESAMIENTO ESTADISTICO DE DATOS**

El proceso estadístico es el conjunto de etapas o fases que deben completarse para realizar una investigación basada en información cuantitativa y obtener unos resultados fieles a la realidad estudiada.

Cuando hablamos de proceso estadístico, estamos hablando de una serie de pasos que es recomendable realizar para obtener unos resultados fieles a la realidad que estudiamos en el estudio estadístico que se pretenda realizar. Esto es necesario, ya que si no realizamos estos pasos podemos obtener conclusiones erróneas y, por ende, tomar malas decisiones. Por ejemplo, imaginemos que tenemos una heladería. Necesitamos saber, en forma aproximada, qué cantidad de helado debemos comprar en función de la cantidad de demanda que tendremos. Así pues, si nos quedamos cortos podrían llegar clientes a los que tendríamos que decirles que en esta heladería no queda helado. Al contrario, si tenemos demasiada cantidad, podría echarse a perder. Por tanto, se hace necesario intentar estimar qué cantidad debemos comprar, o al menos un rango aproximado. Si para calcular ese rango, recolectamos datos que no son representativos (por ejemplo, una heladería situada en otra ciudad con menos afluencia) podríamos equivocarnos.

Así pues, teniendo esto claro, debemos conocer la serie de pasos y detalles que debemos seguir para que los resultados se adecuen a la realidad y tomemos mejores decisiones.

##### **Etapas del proceso estadístico**

Dependiendo del manual visitado o del autor, podríamos ver diferentes etapas con diferentes nombres. En esencia, casi todos los documentos sobre el tema recogen los mismos apartados, solo que unos engloban varias fases en una y otros fragmentan más el proceso.

En nuestro caso, consideramos que el proceso estadístico está formado por:



- a) Planteamiento del problema
- b) Recolección de datos
- c) Organización de datos
- d) Análisis de datos
- e) Interpretación de datos

### 📌 Planteamiento del problema

En el planteamiento del problema se sitúa el eje central sobre el que articular todo lo demás. Esta fase responde a la siguiente pregunta: ¿Qué necesito estudiar y por qué? En ocasiones, por increíble que parezca plantear el problema puede hacernos llegar a la conclusión de que en realidad no necesitamos realizar un estudio estadístico.

Empieza por contextualizar el área o disciplina de estudio donde se enmarca la problemática. Ir de lo general a lo particular. Lograrás tener en cuenta tres aspectos:

1. El espacio de ese contexto, ¿Dónde?: puede ser un espacio real, es decir, un lugar determinado (un país, ciudad, poblado, urbanización, calle, empresa, organización, instituto, etc.) También puede tratarse de un espacio figurado (ciencia, disciplina, corriente de pensamiento, campo de estudio, movimiento literario, etc.)
2. El tiempo ¿Cuándo?: si es una problemática reciente o de larga data.
3. El modo ¿Cómo?: resalta como se presenta la problemática y como se ha estudiado o considerado previamente.

Luego de contextualizar es necesario colocar tu propuesta de análisis, esto incluye la modalidad y las características del estudio que propones en tu trabajo y, si es posible, las soluciones para la resolución de la problemática establecida. Dependiendo de la modalidad del trabajo de grado que escojas, el planteamiento del problema puede contener hipótesis o las interrogantes de la investigación.

Para cerrar el planteamiento del problema expone la necesidad, modalidad y fines de su estudio.

El planteamiento del problema, como has podido notar, es una reflexión ordenada que va dando cuenta de una transición lógica del pensamiento desde el punto de vista metodológico.

### **Recolección de datos**

Una vez hemos planteado el problema debemos recoger los datos. Aquí es importante la metodología. De tal modo que existen diferentes consideraciones. Así pues, debemos establecer el tipo de muestreo, el tamaño de la muestra, el tipo de recolección de datos (por ejemplo, a través de bases de datos o de encuestas personalizadas), en persona, por internet o por teléfono, etc.

Los analistas utilizan una variedad de métodos a fin de recopilar los datos sobre una situación existente, como entrevistas, cuestionarios, inspección de registros (revisión en el sitio) y observación. Cada uno tiene ventajas y desventajas. Generalmente, se utilizan dos o tres para complementar el trabajo de cada una y ayudar a asegurar una investigación completa.

### **La entrevista**

Las entrevistas se utilizan para recabar información en forma verbal, a través de preguntas que propone el analista. Quienes responden pueden ser gerentes o empleados, los cuales son usuarios actuales del sistema existente, usuarios potenciales del sistema propuesto o aquellos que proporcionarán datos o serán afectados por la aplicación propuesta. El analista puede entrevistar al personal en forma individual o en grupos algunos analistas prefieren este método a las otras técnicas que se estudiarán más adelante. Sin embargo, las entrevistas no siempre son la mejor fuente de datos de aplicación.

Dentro de una organización, la entrevista es la técnica más significativa y productiva de que dispone el analista para recabar datos. En otras palabras, la entrevista es un intercambio de información que se efectúa cara a cara. Es un canal de comunicación entre el analista y la organización; sirve para obtener información acerca de las necesidades y la manera de satisfacerlas, así como consejo y comprensión por parte del usuario para toda

idea o método nuevos. Por otra parte, la entrevista ofrece al analista una excelente oportunidad para establecer una corriente de simpatía con el personal usuario, lo cual es fundamental en transcurso del estudio.

### **Preparación de la Entrevista**

1. Determinar la posición que ocupa de la organización el futuro entrevistado, sus responsabilidades básicas, actividades, etc. (Investigación).
2. Preparar las preguntas que van a plantearse, y los documentos necesarios (Organización).
3. Fijar un límite de tiempo y preparar la agenda para la entrevista. (Sicología).
4. Elegir un lugar donde se puede conducir la entrevista con la mayor comodidad (Sicología).
5. Hacer la cita con la debida anticipación (Planeación).

### **Conducción de la Entrevista**

1. Explicar con toda amplitud el propósito y alcance del estudio (Honestidad).
2. Explicar la función propietaria como analista y la función que se espera conferir al entrevistado. (Imparcialidad).
3. Hacer preguntas específicas para obtener respuestas cuantitativas (Hechos).
4. Evitar las preguntas que exijan opiniones interesadas, subjetividad y actitudes similares (habilidad).
5. Evitar el cuchicheo y las frases carentes de sentido (Claridad).
6. Ser cortés y comedido, absteniéndose de emitir juicios de valores. (Objetividad).
7. Conservar el control de la entrevista, evitando las divagaciones y los comentarios al margen de la cuestión.
8. Escuchar atentamente lo que se dice, guardándose de anticiparse a las respuestas (Comunicación).

## **Secuela de la Entrevista**

1. Escribir los resultados (Documentación).
2. Entregar una copia al entrevistado, solicitando su conformación, correcciones o adiciones. (Profesionalismo).
3. Archivar los resultados de la entrevista para referencia y análisis posteriores (Documentación).

## **Recabar datos mediante la Entrevista**

La entrevista es una forma de conversación, no de interrogación, al analizar las características de los sistemas con personal seleccionado cuidadosamente por sus conocimientos sobre el sistema, los analistas pueden conocer datos que no están disponibles en ninguna otra forma.

En las investigaciones de sistema, las formas cualitativas y cuantitativas de la información importantes. La información cualitativa está relacionada con opinión, política y descripciones narrativas de actividades o problemas, mientras que las descripciones cuantitativas tratan con números frecuencia, o cantidades. A menudo las entrevistas pueden ser la mejor fuente de información cualitativas, los otros métodos tiende a ser más útiles en la recabación de datos cuantitativos.

Son valiosas las opiniones, comentarios, ideas o sugerencia en relación a como se podría hacer el trabajo; las entrevistas a veces es la mejor forma para conocer las actividades de las empresas. La entrevista pueden descubrir rápidamente malos entendidos, falsa expectativa o incluso resistenciapotencial para las aplicaciones de desarrollo; más aún, a menudo es más fácil calendarizar una entrevista con los gerentes de alto nivel, que pedirle que llenen cuestionario.

## **Determinación del tipo de Entrevista**

La estructura de la entrevista varia. Si el objetivo de la entrevista radica en adquirir información general, es conveniente elaborar una serie de pregunta sin estructura, con una sesión de preguntas y respuesta libres

Las entrevistas estructuradas utilizan pregunta estandarizada. El formato de respuestas para las preguntas pueden ser abierto o cerrado; las preguntas para respuestas abierta permiten a los entrevistados dar cualquier respuesta que parezca apropiado. Pueden contestar por completo con sus propias palabras. Con las preguntas para respuesta cerradas se proporcionan al usuario un conjunto de respuesta que se pueda seleccionar. Todas las personas que respondes se basan en un mismo conjunto de posible respuestas.

Los analistas también deben dividir el tiempo entre desarrollar preguntas para entrevistas y analizar respuesta. La entrevista no estructurada no requiere menos tiempos de preparación, porque no necesita tener por anticipado las palabras precisas de las preguntas. Analizar las respuestas después de la entrevista lleva más tiempo que con la entrevista estructuradas. El mayor costo radica en la preparación, administración y análisis de las entrevistas estructuradas para pregunta cerradas.

### **Selección de Entrevistados**

Realizar entrevistas toma tiempo; por lo tanto no es posible utilizar este método para recopilar toda la información que se necesite en la investigación; incluso el analista debe verificar los datos recopilados utilizando unos de los otros métodos de recabación de datos. La entrevista se aplican en todos los niveles gerencial y de empleados y dependa de quien pueda proporcionar la mayor parte de la información útil para el estudio los analistas que estudian la administración de inventarios pueden entrevistar a los trabajadores del embarque y de recepción, al personal de almacén y a los supervisores de los diferentes turnos, es decir. Aquellas personas que realmente trabajan en el almacén, también entrevistarán a los gerentes más importante.

### **Realización de Entrevista**

La habilidad del entrevistador es vital para el éxito en la búsqueda de hecho por medio de la entrevista. Las buenas entrevista depende del conocimiento del analista tanto de la preparación del objetivo de una entrevista específica como de las preguntas por realizar a una persona determinada.

El tacto, la imparcialidad e incluso la vestimenta apropiada ayudan a asegurar una entrevista exitosa. La falta de estos factores puede reducir cualquier oportunidad de éxito.

Por ejemplo, analista que trabaja en la aplicación enfocada a la reducción de errores (captado por la gerencia de alto nivel) probablemente no tendría éxito si llegara a una oficina de gerencia de nivel medio con la presentación equivocada, ejemplo "Estamos aquí para resolver su problema".

A través de la entrevista, los analistas deben preguntarse a sí mismo las siguientes preguntas:

- ¿Qué es lo que me está diciendo la persona?
- ¿Por qué me lo está diciendo a mí?
- ¿Qué está olvidando?
- ¿Qué espera esta persona que haga yo?

### **¿Qué es una encuesta?**

Se ha dicho que Estados Unidos ya no es una "sociedad industrial", sino una "sociedad de información". Esto es, nuestros mayores problemas y tareas ya no giran principalmente en la producción de bienes y servicios necesarios para nuestra supervivencia y comodidad.

Nuestra "sociedad", requiere un rápido y preciso flujo de información sobre las preferencias, necesidades y comportamiento de sus miembros. Es en respuesta a esta necesidad crítica de información por el gobierno, el comercio y las instituciones sociales que tanta confianza se pone en las encuestas.

Hoy en día la palabra "encuesta" se usa más frecuentemente para describir un método de obtener información de una muestra de individuos. Esta "muestra" es usualmente sólo una fracción de la población bajo estudio.

Por ejemplo, antes de una elección, una muestra de electores es interrogada para determinar cómo los candidatos y los asuntos son percibidos por el público... un fabricante hace una encuesta al mercado potencial antes de introducir un nuevo producto... una entidad del gobierno comisiona una encuesta para obtener información para evaluar legislación existente o para preparar y proponer nueva legislación.

No tan sólo las encuestas tienen una gran variedad de propósitos, sino que también pueden conducirse de muchas maneras, incluyendo por teléfono, por correo o en persona.

Aún así, todas las encuestas tienen algunas características en común.

A diferencia de un censo, donde todos los miembros de la población son estudiados, las encuestas recogen información de una porción de la población de interés, dependiendo el tamaño de la muestra en el propósito del estudio. En una encuesta *bona fide*, la muestra no es seleccionada caprichosamente o sólo de personas que se ofrecen como voluntarios para participar. La muestra es seleccionada científicamente de manera que cada persona en la población tenga una oportunidad medible de ser seleccionada. De esta manera los resultados pueden ser proyectados con seguridad de la muestra a la población mayor. La información es recogida usando procedimientos estandarizados de manera que a cada individuo se le hacen las mismas preguntas en mas o menos la misma manera. La intención de la encuesta no es describir los individuos particulares quienes, por azar, son parte de la muestra sino obtener un perfil compuesto de la población.

Una "encuesta" recoge información de una "muestra." Una "muestra" es usualmente sólo una porción de la población bajo estudio.

El estándar de la industria para todas las organizaciones respetables que hacen encuestas es que los participantes individuales nunca puedan ser identificados al reportar los hallazgos. Todos los resultados de la encuesta deben presentarse en resúmenes completamente anónimos, tal como tablas y gráficas estadísticas.

### **¿Cuán grande debe ser la muestra?**

El tamaño de muestra requerido en una encuesta depende en parte de la calidad estadística necesaria para los establecer los hallazgos; esto a su vez, está relacionado en cómo esos hallazgos serán usados.

Aún así, no hay una regla simple para el tamaño de muestra que pueda ser usada en todas las encuestas. Mucho de esto depende de los recursos profesionales y fiscales disponibles. Los analistas frecuentemente encuentran que una muestra de tamaño moderado es suficiente estadística y operacionalmente. Por *ejemplo*, las muy conocidas encuestas

nacionales frecuentemente usan cerca de 1,000 personas para obtener información razonable sobre actitudes y opiniones nacionales.

Cuando nos damos cuenta que una muestra apropiadamente seleccionada de sólo 1,000 individuos puede reflejar varias características de la población total, es fácil apreciar el valor de usar encuestas para tomar decisiones informadas en una sociedad compleja como la nuestra. Las encuestas proveen medios rápidos y económicos de determinar la realidad de nuestra economía y sobre los conocimientos, actitudes, creencias, expectativas y comportamientos de las personas.

### **¿Quién lleva a cabo las Encuestas?**

Todos conocemos sobre las encuestas de opinión pública que son reportadas por los medios informativos. Por *ejemplo*, la Encuesta Gallup y la Encuesta Harris emiten informes periódicos describiendo la opinión pública nacional sobre una amplia gama de asuntos corrientes. Encuestas estatales y en las áreas metropolitanas, frecuentemente con el apoyo económico de algún periódico o estación de televisión local, se reportan regularmente en muchos lugares. Las cadenas mayores de radio y televisión, así como revistas nacionales de noticias también llevan a cabo encuestas e informan sus resultados. A pesar de esto, la gran mayoría de las encuestas no son de opinión pública. La mayoría están dirigidas a un propósito administrativo, comercial o científico. La gran variedad de asuntos con los que tratan las encuestas se puede ilustrar con la siguiente lista de usos reales:

- Las cadenas mayores de televisión confían en encuestas que le dicen cuántas y qué tipo de personas ven sus programas.
- Statistics Canadá lleva a cabo encuestas continuas de panel sobre niños (y sus familias) para estudiar sus necesidades educativas y otras.
- Es una buena práctica nunca identificar los participantes individuales. El tamaño de la muestra depende de las metas estadísticas y de los recursos disponibles para la encuesta.
- Los fabricantes de automóviles usan encuestas para determinar cuán satisfechos están las personas con sus autos.



- El Negociado del Censo de los Estados Unidos lleva a cabo encuestas cada mes para obtener información sobre empleo y desempleo en la nación.
- La Agencia para la Política e Investigación sobre Cuidado de Salud de los Estados Unidos auspicia una encuesta periódica para determinar cuanto dinero está gastando la gente en los distintos tipos de cuidado médico.
- Las autoridades de transportación local conducen encuestas para obtener información
- sobre los hábitos de viaje y transportación de las personas.
- Las revistas y revistas profesionales usan encuestas para conocer qué leen sus suscriptores.
- Se llevan a cabo encuestas para conocer quien usa nuestros parques nacionales y
- otras facilidades recreativas.

Las encuestas proveen una fuente importante de conocimiento científico básico. Economistas, psicólogos, profesionales de la salud y sociólogos llevan a cabo encuestas para estudiar materias tales como los patrones de ingreso y gastos en los hogares, las raíces del prejuicio étnico o racial, las implicaciones de los problemas de salud en la vida de las personas, comparando el comportamiento electoral y los efectos sobre la vida familiar de mujeres que trabajan fuera del hogar.

### **Organización de los datos**

Una vez tenemos todos los datos queda unificarlos y organizarlos. Como en todo, necesitamos introducir los datos en programa o plataforma que luego nos permita calcular determinadas métricas y analizar correctamente. Para ello, siempre es conveniente organizar los datos. Es más, a veces necesitaremos recoger datos de diferentes bases de datos que ofrecen formatos de archivos diferentes y será necesario unificarlo todo en el mismo formato.

Vamos a considerar por separado los caso de datos cualitativos y cuantitativos.

### **Organización de los datos cualitativos:**

En este caso la agrupación de los datos es muy sencilla y se hace de acuerdo a las modalidades que presente las variable en estudio. mediante un conteo se determina el número de datos (también llamado frecuencia) correspondiente a las diferentes categorías de la variable. este procedimiento es valido para cualquier cantidad de datos.

### **Organización de los datos cuantitativos:**

para organizar y agrupar datos de tipo cuantitativo discretos o continuos, se utiliza un procedimiento similar, pero más laborioso, al utilizado con los datos cualitativos.

### **Análisis de los datos**

Una vez planteado el problema, recolectados los datos y organizados podemos analizarlos de forma eficaz. Dependiendo del planteamiento del problema, se realizará un tipo de análisis u otro. Por ejemplo, si queremos saber si dos variables son dependientes, podríamos utilizar un análisis de cointegración. Mientras que si lo que queremos estudiar es la dispersión total de un activo financiero, calcularemos el rango estadístico.

El análisis de datos es la ciencia que se encarga de examinar un conjunto de datos con el propósito de sacar conclusiones sobre la información para poder tomar decisiones, o simplemente ampliar los conocimientos sobre diversos temas.

El análisis de datos consiste en someter los datos a la realización de operaciones, esto se hace con la finalidad de obtener conclusiones precisas que nos ayudarán a alcanzar nuestros objetivos, dichas operaciones no pueden definirse previamente ya que la recolección de datos puede revelar ciertas dificultades.

Actualmente, muchas industrias usan el análisis de datos para sacar conclusiones y decidir acciones a implementar. Cabe mencionar que la ciencia también usa el análisis de datos para comprobar o descartar teorías o modelos existentes.

Daniel Burrus, asesor de negocios y orador de temas empresariales y de innovación dice en referencia al análisis de datos: —Mucho de esto ayudará a los humanos a trabajar más, de forma inteligente y rápido,

porque tenemos datos sobre todo lo que ocurre .

### **Usos del análisis de datos:**

El análisis de datos se utiliza en muchas industrias, independientemente del ramo, nos da las bases para tomar o no una decisión o cerciorarnos si una hipótesis es cierta o no.

1. **Mercadotecnia:** el análisis de datos se ha usado principalmente para predecir el comportamiento de los consumidores, incluso para poder calificarlo. Conoce cómo hacer un análisis de datos para tu campaña de marketing.
2. **Recursos Humanos:** el análisis de datos también es muy útil dentro de las empresas para mantener un buen clima laboral, y fuera de ella, calificando empleados potenciales.
3. **Académicos:** Al igual que las empresas el análisis de datos también está presente en la educación, sirve para seleccionar a los alumnos de nuevo ingreso y para medir el rendimiento de los estudiantes.

### **Técnicas de análisis de datos:**

Si queremos datos útiles, debemos analizarlos. Para ello debemos recurrir a diversas técnicas que dependen del tipo de información que se esté recopilando, por lo que es importante tener definida la técnica a utilizar antes de implementarla.

- **Análisis de datos cualitativo:** Los datos cualitativos se presentan de manera verbal (en ocasiones en gráficas). Se basa en la interpretación. Las formas más comunes de obtener esta información es a través de entrevistas abiertas, grupos de discusión y grupos de observación, donde los investigadores generalmente analizan patrones en las observaciones durante toda la fase de recolección de datos.
- **Análisis de datos cuantitativos:** Los datos cuantitativos se presentan en forma numérica. Se basa en resultados tangibles.

El análisis de datos se centra en llegar a una conclusión basada únicamente en lo que ya es conocido por el investigador. La forma en que recopila sus datos debe relacionarse con la forma en que está planeando analizarla y utilizarla, también hay que asegurarse de recopilar información precisa en la que puedas confiar, para ello existen muchas técnicas de recolección de datos.

La técnica más usada por los expertos son las encuestas online, ya que puede traer grandes beneficios como la reducción de tiempo y dinero.

## Ventajas del análisis de datos

- Capacidad para tomar decisiones de negocios más rápidas e informadas, respaldadas por hechos.
- Ayuda a las empresas a identificar problemas de rendimiento que requieren algún tipo de acción.
- Comprensión más profunda de los requisitos de los clientes, lo que, a su vez, crea mejores relaciones comerciales.
- Mayor conciencia del riesgo, permitiendo la implementación de medidas preventivas.
- Puede verse de forma visual, lo que permite tomar decisiones más rápidas y mejores.
- Puede proporcionar a una empresa una ventaja sobre sus competidores.
- Mejor conocimiento del desempeño financiero del negocio.
- Se ha demostrado que reduce los costos y, por lo tanto, aumenta los beneficios.

### Tipos de análisis de datos

	Tipos de datos	Análisis	Ejemplos
<b>Cualitativo</b>	Se centra en las opiniones, actitudes y creencias.	Preguntas y respuestas a preguntas como: ¿Por qué? ¿Cómo?	Paneles en donde se da una discusión y se entrevista a consumidores sobre lo que les agrada o no del lugar.
<b>Cuantitativo</b>	Se centra en los datos duros e información que pueda contabilizarse.	Se obtiene mediante preguntas similares a: ¿Cuántos? ¿Quién? ¿Con qué frecuencia? ¿Dónde?	Encuestas enfocadas a medir las ventas, tendencias, reportes o percepciones.

## **Pasos para hacer un análisis de datos**

### **Paso 1: Define tus preguntas**

Comienza seleccionando las preguntas correctas. Las preguntas deben ser medibles, claras y concisas. Diseña sus preguntas para calificar o descalificar posibles soluciones a su problema u oportunidad específicos.

### **Paso 2: Establece prioridades de medición**

Este paso se divide en dos sub-pasos:

- A) Decide qué medir: Analiza qué tipo de datos necesitas.
- B) Decidir cómo medirlo: Pensar en cómo medir sus datos es igual de importante, especialmente antes de la fase de recolección de datos, porque su proceso de medición respalda o desacredita su análisis más adelante.

### **Paso 3: Recolecta datos**

Con la pregunta claramente definida y sus prioridades de medición establecidas, ahora es el momento de recopilar sus datos. A medida que recopiles y organices los datos, recuerda tener en cuenta estos puntos importantes:

Antes de recopilar nuevos datos, determina qué información podría recopilarse de las bases de datos o fuentes existentes.

Determina de antemano un sistema de almacenamiento y asignación de nombres de archivos para ayudar a todos los miembros del equipo a colaborar. Este proceso ahorra tiempo y evita que los miembros del equipo recopilen la misma información dos veces.

Si necesita recopilar datos mediante encuestas, observación o entrevistas, desarrolla con anticipación un cuestionario para asegurar la consistencia y ahorrar tiempo.

Mantén los datos recopilados organizados en un registro con las fechas de recopilación y agrega cualquier nota de origen a medida que avanza.

Quizá te interese leer: ¿Qué es la investigación primaria y secundaria?

#### **Paso 4: Analiza los datos**

Una vez que haya recopilado los datos correctos para responder a su pregunta del Paso 1, es el momento de realizar un análisis más profundo de la información. Encuentra relaciones, tendencias, ordena y filtra tu información de acuerdo a las variables. A medida que haces un análisis de los datos encontrarás que tienes los datos exactos que necesitas.

Te recomendamos leer: [¿Cómo analizar los datos de una investigación?](#)

#### **Paso 5: Interpretar los resultados**

Después de analizar los datos y posiblemente realizar más investigaciones, finalmente es tiempo de interpretar los resultados. Hazte estas preguntas clave:

- ¿Responden los datos a tu pregunta original? ¿Cómo?
- ¿Los datos te ayudan a defender cualquier objeción? ¿Cómo?
- ¿Hay alguna limitación en las conclusiones, algún ángulo que no hayas considerado?

Si tu interpretación de los datos se sostiene bajo todas estas preguntas y consideraciones, entonces es probable que hayas llegado a una conclusión productiva. El único paso restante es utilizar los resultados del proceso de análisis de datos para decidir cómo vas a actuar.

Con estos cinco pasos en tu proceso de análisis de datos, tomarás mejores decisiones para tu negocio ya que tus elecciones están respaldadas por datos que han sido robustamente recopilados y analizados.

#### **Interpretación de los datos**

Por último, pero no por ello menos importante, tenemos la interpretación de los datos. De nada sirve realizar todas las fases del proceso estadístico correctamente si al final la interpretación es errónea. Esto es debido a que si la interpretación es errónea, entonces las decisiones tendrán un efecto no deseado. Por ejemplo, imaginemos que realizamos un estudio sobre la variabilidad de las ventas de una empresa. Si una vez obtenemos los resultados resulta que hay mucha dispersión, conviene reducirla y nosotros interpretamos que no, esto podrá afectar negativamente a la empresa

## **Interpretación cuantitativa de datos**

Si la interpretación de los datos cuantitativos se pudiera resumir en una palabra (y realmente no puede) esa palabra sería "numérica".

Hay pocas certezas en lo que respecta al análisis de datos, pero puede estar seguro de que si la investigación es interesante no tiene números involucrados, no es una investigación cuantitativa.

El análisis cuantitativo se refiere a un conjunto de procesos mediante los cuales se analizan los datos numéricos. En la mayoría de los casos, implica el uso de modelos estadísticos como la desviación estándar, la media y la mediana.

## **Beneficios de la Interpretación de datos**

¿Cuáles son los beneficios de la interpretación de datos? ¿Por qué todas las industrias participan en la investigación y análisis de datos? Es una pregunta básica, pero que a menudo no recibe la atención adecuada.

El objetivo de la recopilación e interpretación es adquirir información útil y utilizable y tomar las decisiones más informadas posibles.

Desde empresas, hasta educación superior, hasta recién casados que investigan su primer hogar, la recopilación e interpretación de datos puede proporcionar beneficios ilimitados para una amplia gama de instituciones y particulares.

El análisis e interpretación de los datos, independientemente del método y del estado cualitativo / cuantitativo, pueden incluir las siguientes características:

- Identificación de datos y explicación
- Comparación y contraste de datos
- Identificación de datos atípicos
- Predicciones futuras

El análisis e interpretación de datos, al final, ayuda a mejorar los procesos e identificar problemas.

## **Beneficios Comerciales de la interpretación de datos**

**Toma de decisiones informada:** una decisión es tan buena como el conocimiento que la formó.

La toma de decisiones informadas sobre los datos tiene el potencial de diferenciar a los líderes de la industria del resto del mercado.

Los estudios han demostrado que las empresas en el tercio superior de sus industrias son, en promedio, un 5% más productivas y un 6% más rentables cuando implementan procesos informa

Las acciones más decisivas surgirán solo después de que se haya identificado un problema o se haya definido un objetivo.

El análisis de los datos debe incluir la identificación, el desarrollo y la recopilación de datos seguidos de la comunicación de datos.

**Identificación de tendencias:** los conocimientos de datos proporcionan conocimiento y el conocimiento es poder.

Los conocimientos obtenidos de los análisis de datos del mercado y del consumidor tienen la capacidad de servir para inferir información de segmentos de mercados similares.

**Eficiencia en costos:** la implementación adecuada de los procesos de análisis de datos puede proporcionar a las empresas profundas ventajas de costos dentro de sus industrias. Un reciente estudio de datos realizado por Deloitte demuestra vívidamente esto al encontrar que el ROI del análisis de datos se basa en reducciones de costos eficientes.

A menudo, este beneficio se pasa por alto porque, en general, se considera que ganar dinero es más "sexy" que ahorrar dinero.

Sin embargo, los análisis de datos sólidos tienen la capacidad de alertar a la gerencia de las oportunidades de reducción de costos sin ningún esfuerzo significativo por parte del capital humano.

## **BIBLIOGRAFIA:**

1.-ALEA, V. et al. (2006) *Estadística Aplicada a les Ciències Econòmiques i Socials*. Barcelona: Edicions McGraw-Hill EUB.



- 2.-CANAVOS, G. (2008) *Probabilidad y Estadística. Aplicaciones y Métodos*. México: McGraw-Hill.
- 3.-DURAPEIRÓ, J. M. y LÓPEZCUÑAT, J.M. (2006) *Fundamentos de Estadística. Estadística Descriptiva y Modelos Probabilísticos para la Inferencia*. Madrid: Ariel Editorial.
- 4.-ESCUDER, R. y SANTIAGO, J. (2010) *Estadística aplicada. Economía y Ciencias Sociales*. Valencia: Tirant lo Blanch.
- 5.-FERNÁNDEZ CUESTA, C., y FUENTES GARCÍA, F. (2015) *Curso de Estadística Descriptiva Teoría y Práctica*. Madrid: Ariel.
- 6.<http://paginas.facmed.unam.mx/deptos/sp/wpcontent/uploads/2015/11/03REYNAGA1.pdf>

## 1.2 DISTRIBUCIONES DE FRECUENCIAS

Las tablas de distribución de frecuencias se utilizan cuando se recolectan datos, con ellas se pueden representar los datos de manera que es más fácil analizarlos.

Se pueden elaborar tablas de distribución de frecuencias para datos no agrupados y para datos agrupados. Estas últimas se utilizan cuando se tienen muchos datos.

Para elaborar tablas de distribuciones de frecuencias se debe tener en cuenta lo siguiente: Cuando hay muchos datos se agrupan en clases. Esto consiste en agrupar los datos en una distribución de frecuencias, que puede definirse como una ordenación o arreglo de datos en clases o categorías que muestran para cada una de ellas, el número de elementos que contiene, denominada frecuencia



Este diagrama facilita determinar la cantidad de veces que se repite un dato y los valores de los datos con el fin de escribirlos de manera ordenada en la tabla.

Para construir la tabla de datos no agrupados se debe calcular primero lo siguiente:

**Número de clases**

$$k=1+3,322\log(n)=1+3,322\log(50)=6,64\approx 7$$

**Rango**

$$R=x_n-x_1=72-53=19$$

**Amplitud de clase**

$$l=R/k=19/7=2,71\approx 3$$

**Punto medio:**  $m_i$  es el valor central de la clase. Se obtiene calculando el promedio de los límites reales, sumando al límite real inferior el límite real superior y dividiendo por dos.

**Frecuencia absoluta.** Se define como el número de elementos u observaciones pertenecientes a una misma clase.

**Frecuencia relativa:** Se obtiene dividiendo la frecuencia absoluta por el número total de observaciones. Indica la importancia relativa de la clase.

**Frecuencias acumuladas:** Es la suma de las frecuencias absolutas o relativas en sentido ascendente o descendente según se quieran acumular —hacia arriba o —hacia abajo

Al construir la tabla de datos agrupados con la información del ejemplo, se tiene:

## Tabla de datos agrupados

	Punto medio	Frecuencia absoluta	Frecuencia absoluta acumulada	Frecuencia relativa	Frecuencia relativa acumulada
<b>Pesos</b>	$m_i$	$f_i$	$F_i$	$fr_i$	$Fr_i$
<b>(Kg)</b>					
53 - 55	54	2	2	4,00%	4,00%
56 - 58	57	5	7	10,00%	14,00%
59 - 61	60	9	16	18,00%	32,00%
62 - 64	63	15	31	30,00%	62,00%
65 - 67	66	12	43	24,00%	86,00%
68 - 70	69	5	48	10,00%	96,00%
71 - 73	72	2	50	4,00%	100,00%
		50		100,00%	

## BIBLIOGRAFIA:

1. FREEDMAN, D., et al. (2001) *Estadística*. Barcelona: A. Bosch Ed.
2. FREEDMAN, D., et al. (2015) *Estadística*. Barcelona: A. Bosch Ed.
3. FREIXA, M., et al. (2012) *Análisis exploratorio de datos: Nuevas técnicas estadísticas*. Barcelona: PPU.
4. GUJARATI, D. (2007) *Econometría Básica*. Bogotá: McGraw-Hill.
5. KMENTA, J (2011) *Elementos de Econometría*. Barcelona: Vicens Universidad.

## 1.3 PRESENTACION GRAFICA

Toda investigación de índole científico se apoya y base en un conjunto de datos debidamente analizado e interpretado. Para llegar a un punto en que podamos extraer relaciones de causalidad o de correlación es necesario observar múltiples observaciones de manera que se pueda falsear y comprobar la existencia de la misma relación en diferentes casos o en el mismo sujeto a través del tiempo. Y una vez hechas dichas observaciones hace falta tener en cuenta aspectos como la frecuencia, la media, la moda o la dispersión de los datos obtenidos.

Con la finalidad de facilitar la comprensión y el análisis tanto por parte de los mismos investigadores como de cara a mostrar la variabilidad de los datos y de donde salen las conclusiones al resto del mundo, es de gran utilidad emplear elementos visuales de fácil interpretación: las gráficas o gráficos.

En función de lo que queramos mostrar, podemos emplear diversos tipos de gráficas. En este artículo veremos diferentes tipos de gráficas que se emplean en investigación a partir del uso de la estadística.

#### El gráfico

A un nivel estadístico y matemático, denominados gráfica a aquella representación visual a partir de la cual pueden representarse e interpretarse valores generalmente numéricos. De entre las múltiples informaciones extraíbles de la observación de la gráfica podemos encontrar la existencia de relación entre variables y el grado en que se da, las frecuencias o la proporción de aparición de determinadas valores. Esta representación visual sirve de apoyo a la hora de mostrar y comprender de manera sintetizada los datos recabados durante la investigación, de manera que puede tanto los investigadores que llevan a cabo el análisis como otros puedan comprender los resultados y resulte sencillo utilizarlo como referencia, como información a tener en cuenta o como punto de contraste ante la realización de nuevas investigaciones y metaanálisis.

#### Tipos de gráficas

Existen muy diversos tipos de gráficas, generalmente aplicándose unas u otras en función de lo que se pretenda representar o simplemente de las preferencias del autor. A continuación indicamos algunas de las más conocidas y comunes.

#### Gráfico de barras

El más conocido y utilizado de todos los tipos de gráficos es el gráfico o diagrama de barras. En éste, se presentan los datos en forma de barras contenidas en dos ejes cartesianos (coordenada y abscisa) que indican los diferentes valores. **El aspecto visual que nos indica los datos es la longitud de dichas barras**, no siendo importante su grosor.

Generalmente se emplea para representar la frecuencia de diferentes condiciones o variables discretas (por ejemplo la frecuencia de los diferentes colores del iris en una muestra determinada, que solo pueden ser unos valores concretos). Únicamente se observa una variable en las abscisas, y las frecuencias en las coordenadas.

### **Gráfico circular o por sectores**

El también muy habitual gráfico en forma de —quesito , en este caso la representación de los datos se lleva a cabo mediante la división de un círculo en tantas partes como valores de la variable investigada y teniendo cada parte **un tamaño proporcional a su frecuencia dentro del total de los datos**. Cada sector va a representar un valor de la variable con la que se trabaja.

Este tipo de gráfico o diagrama es habitual cuando se está mostrando la proporción de casos dentro del total, utilizando para representarlo valores percentuales (el porcentaje de cada valor).

### **Histograma**

Aunque a simple vista muy semejante al gráfico de barras, el histograma es uno de los tipos de gráfica que a nivel estadístico resulta más importante y fiable. En esta ocasión, también se utilizan barras para indicar a través de ejes cartesianos la frecuencia de determinados valores, pero en vez de limitarse a establecer la frecuencia de un valor concreto de la variable evaluada refleja todo un intervalo. Se observa pues un rango de valores, que además **podrían llegar a reflejar intervalos de diferentes longitudes**.

Ello permite observar no solo la frecuencia sino también la dispersión de un continuo de valores, lo que a su vez puede ayudar a inferir la probabilidad. Generalmente se utiliza ante variables continuas, como el tiempo.

### **Gráfico de líneas**

En este tipo de gráfico se emplean líneas para **delimitar el valor de una variable dependiente respecto a otra independiente**. También puede usarse para comparar los valores de una misma variable o de diferentes investigaciones utilizando el mismo gráfico (usando diferentes líneas). Es usual que se emplee para observar la evolución de una variable a través del tiempo

### **Gráfico de dispersión**

El gráfico de dispersión o gráfico xy es un tipo de gráfico en el cual mediante los ejes cartesianos se representa en forma de puntos todos los datos obtenidos mediante la observación. **Los ejes x e y muestran cada uno los valores de una variable dependiente y otra independiente** o dos variables de la que se esté observando si presentan algún tipo de relación.

Los puntos representados el valor reflejado en cada observación, lo que a nivel visual dejará ver una nube de puntos a través de los cuales podemos observar el nivel de dispersión de los datos.

Se puede observar si existe o no una relación entre las variables mediante el cálculo. Es el procedimiento que se suele usar, por ejemplo, para establecer la existencia de rectas de regresión lineal que permita determinar si hay relación entre variables e incluso el tipo de relación existente.

### **Gráfico de caja y bigotes**

Los gráficos de caja son uno de los tipos de gráficas que tienden a utilizarse de cara a observar la dispersión de los datos y cómo éstos agrupan sus valores. Se parte del cálculo de los cuartiles, los cuales son los valores que **permiten dividir los datos en cuatro partes iguales**. Así, podemos encontrar un total de tres cuartiles (el segundo de los cuales se corresponderían con la mediana de los datos) que van a configurar la —caja— en cuestión. Los llamados bigotes serían la representación gráfica de los valores extremos.

Este gráfico **es útil a la hora de evaluar intervalos**, así como de observar el nivel de dispersión de los datos a partir de los valores de los cuartiles y los valores extremos.

### **Gráfico de áreas**

En este tipo de gráfico se observa, de manera semejante lo que ocurre con los gráficos de líneas, la relación entre variable dependiente e independiente. Inicialmente se hace una línea que une los puntos que marcan los diferentes valores de la variable medida, pero también se incluye todo lo situado por debajo: este tipo de gráfica nos permite ver la acumulación (un punto determinado incluye a los situados por debajo). A través de él se pueden medir y comparar los valores de diferentes muestras (por ejemplo, comparar los resultados obtenidos por dos personas, compañías, países, por dos registros de un mismo valor....). Los diferentes resultados pueden apilarse, observándose fácilmente las diferencias entre las diversas muestras.

### **Pictograma**

Se entiende por pictograma a un gráfico en el que, en vez de representar los datos a partir de elementos abstractos como barras o círculos, **se emplean elementos propios del tema que se está investigando**. De este modo se hace más visual. Sin embargo, su funcionamiento es semejante al del gráfico de barras, representando frecuencias de la misma manera

### **Cartograma**

Este gráfico resulta de utilidad en el terreno de la epidemiología, indicando las zonas o áreas geográficas en las que aparece con mayor o menor frecuencia un determinado valor de una variable. Las frecuencias o rangos de frecuencias se indican mediante el uso del color (requiriéndose una leyenda para comprenderse) o el tamaño

### **BIBLIOGRAFIA:**

Martínez-González, M.A.; Faulin, F.J. y Sánchez, A. (2006). Bioestadística amigable, 2ª ed. Diaz de Santos, Madrid.

## **1.4 MEDIDAS DE TENDENCIA CENTRAL**

Las características globales de un conjunto de datos estadísticos pueden resumirse mediante una serie de cantidades numéricas representativas llamadas parámetros estadísticos. Entre ellas, las medidas de tendencia central, como la media aritmética, la moda o la mediana, ayudan a conocer de forma aproximada el comportamiento de una distribución estadística.

### Medidas de centralización

---

Se llama medidas de posición, tendencia central o centralización a unos valores numéricos en torno a los cuales se agrupan, en mayor o menor medida, los valores de una variable estadística. Estas medidas se conocen también como promedios.

Para que un valor pueda ser considerado promedio, debe cumplirse que esté situado entre el menor y el mayor de la serie y que su cálculo y utilización resulten sencillos en términos matemáticos.



Se distinguen dos clases principales de valores promedio:

- Las medidas de posición centrales: medias (aritmética, geométrica, cuadrática, ponderada), mediana y moda.
- Las medidas de posición no centrales: entre las que destacan especialmente los cuantiles.



Las medidas de centralización son parámetros representativos de distribuciones de frecuencia como las que ilustra la imagen.

#### Media aritmética

Se define media aritmética de una serie de valores como el resultado producido al sumar todos ellos y dividir la suma por el número total de valores. La media aritmética se expresa como .

Dada una variable  $\bar{x}$  que toma los valores  $x_1, x_2, \dots, x_n$ , con frecuencias absolutas simbolizadas por  $f_1, f_2, \dots, f_n$ , la media aritmética de todos estos valores vendrá dada por:

$$\bar{x} = \frac{f_1 x_1 + f_2 x_2 + \dots + f_n x_n}{f_1 + f_2 + \dots + f_n} = \frac{\sum f_i x_i}{\sum f_i} \text{ con } i = 1, 2, \dots, n$$

#### Media ponderada

En algunas series estadísticas, no todos los valores tienen la misma importancia. Entonces, para calcular la media se ponderan dichos valores según su peso, con lo que se obtiene una media ponderada.

Si se tiene una variable con valores  $x_1, x_2, \dots, x_n$ , a los que se asigna un peso mediante valores numéricos  $p_1, p_2, \dots, p_n$ , la media ponderada se calculará como sigue:

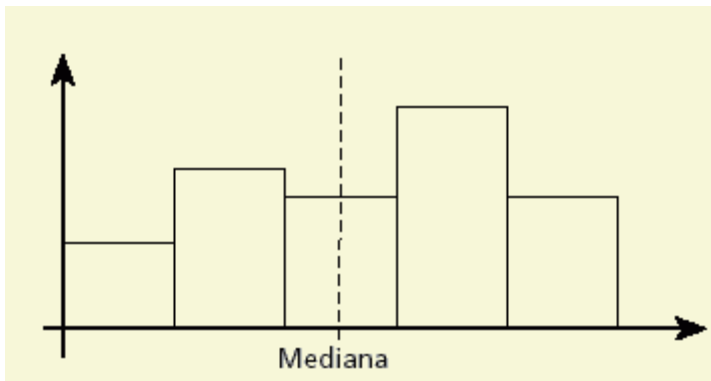
$$\bar{x} = \frac{p_1 x_1 + p_2 x_2 + \dots + p_n x_n}{p_1 + p_2 + \dots + p_n} = \frac{\sum p_i x_i}{\sum p_i} \text{ con } i = 1, 2, \dots, n$$

## Mediana

La media aritmética no siempre es representativa de una serie estadística. Para complementarla, se utiliza un valor numérico conocido como mediana o valor central.

Dado un conjunto de valores ordenados, su mediana se define como un valor numérico tal que se encuentra en el centro de la serie, con igual número de valores superiores a él que inferiores. Normalmente, la mediana se expresa como Me.

La mediana es única para cada grupo de valores. Cuando el número de valores ordenados (de mayor a menor, o de menor a mayor) de la serie es impar, la mediana corresponderá al valor que ocupe la posición  $(n + 1)/2$  de la serie. Si el número de valores es par, ninguno de ellos ocupará la posición central. Entonces, se tomará como mediana la media aritmética entre los dos valores centrales.



Determinación de la mediana de una serie de valores. Moda

En una serie de valores a los que se asocia una frecuencia, se define moda como el valor de la variable que posee una frecuencia mayor que los restantes. La moda se simboliza normalmente por Mo.

Un grupo de valores puede tener varias modas. Una serie de valores con sólo una moda se denomina unimodal; si tiene dos modas, es bimodal, y así sucesivamente.

## BIBLIOGRAFIA:

1.-MARTÍN PLIEGO, F. (1994) *Introducción a la Estadística Económica y Empresarial*. (Teoría y Práctica) Madrid: AC.

2.-MARTÍN PLIEGO, F. y RUIZ-MAYA, L. (1995) *Estadística I: Probabilidad*. Madrid: AC. 3.-MARTÍN

PLIEGO, F. y RUIZ-MAYA, L. (1995) *Estadística II: Inferencia*. Madrid: AC.

## 1.5- MEDIDAS DE DISPERSION

Las medidas de tendencia central ofrecen una idea aproximada del comportamiento de una serie estadística. No obstante, no resultan suficientes para expresar sus características: una misma media puede provenir de valores cercanos a la misma o resultar de la confluencia de datos estadísticos enormemente dispares. Para conocer en que grado las medidas de tendencia central son representativas de la serie, se han de complementar con medidas de dispersión como la varianza o la desviación típica.

### Concentración y dispersión

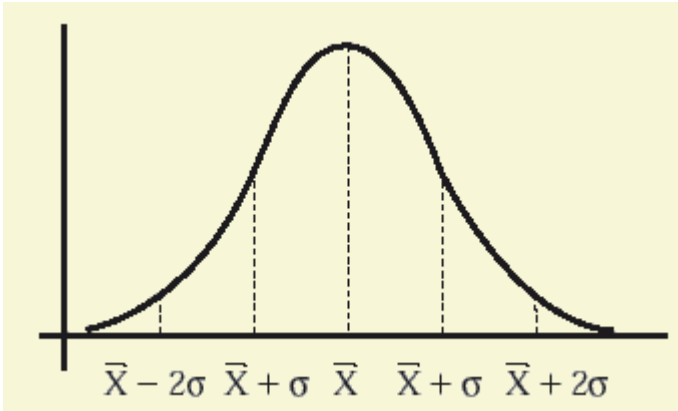
---

Las medidas de centralización ayudan a determinar el «centro de gravedad» de una distribución estadística. Para describir el comportamiento general de la serie se necesita, sin embargo, una información complementaria para saber si los datos están dispersos o agrupados.

Así, las medidas de dispersión pueden definirse como los valores numéricos cuyo objeto es analizar el grado de separación de los valores de una serie estadística con respecto a las medidas de tendencia central consideradas.

Las medidas de dispersión son de dos tipos:

- Medidas de dispersión absoluta: como recorrido, desviación media, varianza y desviación típica, que se usan en los análisis estadísticos generales.
- Medidas de dispersión relativa: que determinan la dispersión de la distribución estadística independientemente de las unidades en que se exprese la variable. Se trata de parámetros más técnicos y utilizados en estudios específicos, y entre ellas se encuentran los coeficientes de apertura, el recorrido relativo, el coeficiente de variación (índice de dispersión de Pearson) y el índice de dispersión mediana.



La distribución normal, o campana de Gauss, es una función simétrica (con la media aritmética en el centro de la serie) con un grado de dispersión bajo (la mayoría de los valores están comprendidos dentro del valor de la desviación típica).

#### Recorrido

La medida de dispersión más inmediata es el recorrido de la distribución estadística, también llamado rango o amplitud. Dada una serie de valores  $x_1, x_2, \dots, x_n$ , su recorrido es la diferencia aritmética entre el máximo y el mínimo de estos valores:

$$Re = x_i (\text{máx}) - x_i (\text{mín}), \text{ siendo } i = 1, 2, \dots, n.$$

#### Desviación media

Como medida de dispersión más frecuentemente utilizada, la desviación media se define como la media aritmética de los valores absolutos de la desviación de cada valor de la variable con respecto a la media. Su formulación matemática es la siguiente:

$$DM = \frac{f_1 |x_1 - \bar{x}| + f_2 |x_2 - \bar{x}| + \dots + f_n |x_n - \bar{x}|}{f_1 + f_2 + \dots + f_n} = \frac{\sum f_i |x_i - \bar{x}|}{\sum f_i} \text{ con } i = 1, 2, 3, \dots, n.$$

#### Varianza y desviación típica

La desviación media no siempre suministra una idea clara del grado de separación entre los valores de una variable estadística. Para estudios científicos, se prefiere utilizar una pareja de parámetros relacionados que se conocen como varianza y desviación típica.

La varianza se define como el cociente entre la suma de los cuadrados de las desviaciones de los valores de la variable y el número de datos del estudio. Matemáticamente, se expresa como:

$$V = \frac{f_1(x_1 - \bar{x})^2 + f_2(x_2 - \bar{x})^2 + \dots + f_n(x_n - \bar{x})^2}{f_1 + f_2 + \dots + f_n} =$$
$$= \frac{\sum f_i(x_i - \bar{x})^2}{\sum f_i} \text{ con } i = 1, 2, 3, \dots, n.$$

Por su parte, la desviación típica, simbolizada por  $s$ , se define sencillamente como la raíz cuadrada de la varianza:

$$\sigma = \sqrt{V}$$

Por lo tanto, se tiene que:

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

La varianza y la desviación típica, cada una con su respectivo valor, se usan indistintamente en los estudios estadísticos.

## **BIBLIOGRAFIA**

1.-MARTÍN-GUZMÁN, P. y MARTÍN PLIEGO, F. (1985) *Curso Básico de Estadística Económica*. Madrid: AC.

2.-MENDENHALL, W., et al. (1994) *Estadística Matemática con Aplicaciones*. México: Grupo Editorial Iberoamérica.

3.-MONTIEL, A.M., RIUS, F. y BARÓN, F.J. (1997) *Elementos Básicos de Estadística Económica y Empresarial*. Madrid: Prentice Hall.

### 1.6- TEOREMA DE CHEBYSHEV

El **teorema de Chebyshev** (o desigualdad de Chebyshev) es uno de los resultados clásicos más importantes de la teoría de la probabilidad. Permite estimar la probabilidad de un evento descrito en términos de una variable aleatoria  $X$ , al proveernos de una cota que no depende de la distribución de la variable aleatoria sino de la varianza de  $X$ .

El teorema recibe el nombre en honor al matemático ruso Pafnuty Chebyshev (también escrito como Chebychev o Tchebycheff) quien, a pesar de no ser el primero en enunciar dicho teorema, fue el primero en dar una demostración en el año 1867.

En el estudio de la teoría de la probabilidad ocurre que si se conoce la función de distribución de una variable aleatoria  $X$ , se puede calcular su valor esperado —o esperanza matemática  $E(X)$ — y su varianza  $Var(X)$ , siempre y cuando dichas cantidades existan. Sin embargo, el recíproco no es necesariamente cierto.

Es decir, conociendo  $E(X)$  y  $Var(X)$  no necesariamente se puede obtener la función de distribución de  $X$ , por lo cual cantidades como  $P(|X| > k)$  para algún  $k > 0$ , son muy difíciles de obtener. Pero gracias a la desigualdad de Chebyshev es posible hacer una estimación de la probabilidad de la variable aleatoria.

El teorema de Chebyshev nos dice que si tenemos una variable aleatoria  $X$  sobre un espacio muestral  $S$  con una función de probabilidad  $p$ , y si  $k > 0$ , entonces:

$$P(|X(s) - E(X)| \geq k) \leq \frac{Var(X)}{k^2}$$

Aplicaciones y ejemplos

Dentro de las muchas aplicaciones que posee el teorema de Chebyshev se pueden mencionar las siguientes:

Acotamiento de probabilidades

Esta es la aplicación más común y se utiliza para dar una cota superior para  $P(|X-E(X)| \geq k)$  donde  $k > 0$ , solo con la varianza y la esperanza de la variable aleatoria  $X$ , sin conocer la función de probabilidad.

*Ejemplo 1*

Supongamos que el número de productos fabricados en una empresa durante una semana es una variable aleatoria con promedio de 50.

Si se sabe que la varianza de una semana de producción es igual a 25, entonces ¿qué podemos decir acerca de la probabilidad de que en esta semana la producción difiera en más de 10 a la media?

**Solución**

Aplicando la desigualdad de Chebyshev tenemos que:

$$P(|X(s) - 50| \geq 10) \leq \frac{Var(X)}{10^2} = \frac{1}{4}$$

De esto podemos obtener que la probabilidad de que en la semana de producción el número de artículos exceda en más de 10 a la media es a lo más 1/4.

## Demostración de los teoremas límites

La desigualdad de Chebyshev juega un papel importante en la demostración de los teoremas límites más importantes. Como ejemplo tenemos los siguientes:

### *Ley débil de los grandes números*

Esta ley establece que dada una sucesión  $X_1, X_2, \dots, X_n, \dots$  de variables aleatorias independientes con la misma distribución promedio  $E(X_i) = \mu$  y varianza  $\text{Var}(X) = \sigma^2$ , y una muestra media conocida de:

$$S_n = (X_1 + X_2 + \dots + X_n)/n$$

Entonces para  $k > 0$  se tiene que:

$$\lim_{n \rightarrow \infty} P(|S_n - \mu| \geq k) = 0$$

O, de manera equivalente:

$$\lim_{n \rightarrow \infty} P(|S_n - \mu| < k) = 1$$

### **Demostración**

Primero notemos lo siguiente:

$$E(S_n) = \frac{E(X_1) + E(X_2) + \dots + E(X_n)}{n} = \frac{n\mu}{n} = \mu$$



Como  $X_1, X_2, \dots, X_n$  son independientes, se deduce que:

$$\text{Var}(X_1 + X_2 + \dots + X_n) = \text{Var}(X_1) + \text{Var}(X_2) + \dots + \text{Var}(X_n)$$

Por lo tanto, es posible afirmar lo siguiente:

$$\text{Var}(S_n) = \text{Var}\left(\frac{X_1 + X_2 + \dots + X_n}{n}\right) = \frac{\text{Var}(X_1 + X_2 + \dots + X_n)}{n^2}$$

$$\frac{\text{Var}(X_1 + X_2 + \dots + X_n)}{n^2} = \frac{n\sigma^2}{n^2} = \frac{\sigma^2}{n}$$

Luego, usando el teorema de Chebyshev se tiene que:

$$P(|S_n - \mu| \geq k) \leq \frac{\sigma^2}{nk^2}$$

Finalmente, el teorema resulta del hecho de que el límite a la derecha es cero cuando  $n$  tiende a infinito.

Cabe resaltar que esta prueba se hizo solo para el caso en el que exista la varianza de  $X_i$ ; es decir, que no diverge. Así observamos que el teorema siempre es verdadero si  $E(X_i)$  existe.

#### *Teorema límite de Chebyshev*

Si  $X_1, X_2, \dots, X_n, \dots$  es una sucesión de variables aleatorias independientes tal que existe algún  $C < \infty$ , tal que  $\text{Var}(X_n) \leq C$  para todo  $n$  natural, entonces para cualquier  $k > 0$ :

$$\lim_{n \rightarrow \infty} P\left(|S_n - \frac{1}{n} \sum_{i=1}^n E(X_i)| < k\right) = 1$$

### **Demostración**

Como la sucesión de varianzas es uniformemente acotada, tenemos que  $\text{Var}(S_n) \leq C/n$ , para todo  $n$  natural.

Pero sabemos que:

$$P\left(|S_n - \frac{1}{n} \sum_{i=1}^n E(X_i)| < k\right) \geq 1 - \frac{C}{n k^2}$$

Haciendo tender  $n$  hacia infinito, resulta lo siguiente:

$$\lim_{n \rightarrow \infty} P\left(|S_n - \frac{1}{n} \sum_{i=1}^n E(X_i)| < k\right) \geq 1$$

Como una probabilidad no puede exceder el valor de 1, se obtiene el resultado deseado. Como consecuencia de este teorema podríamos mencionar el caso particular de Bernoulli.

Si un experimento se repite  $n$  veces de forma independiente con dos resultados posibles (fracaso y éxito), donde  $p$  es la probabilidad de éxito en cada experimento y  $X$  es la variable aleatoria que representa el número de éxitos obtenidos, entonces para cada  $k > 0$  se tiene que:

$$\lim_{n \rightarrow \infty} P\left(\left|\frac{X_n}{n} - p\right| < k\right) = 1$$

Tamaño de muestra

En términos de la varianza, la desigualdad de Chebyshev nos permite encontrar un tamaño de muestra  $n$  que es suficiente para garantizar que la probabilidad de que  $|S_n - \mu| \geq k$  ocurra sea tan pequeña como se desee, lo cual permite tener una aproximación a la media.

De manera precisa, sea  $X_1, X_2, \dots, X_n$  una muestra de variables aleatorias independientes de tamaño  $n$  y supongamos que  $E(X_i) =$

$\mu$  y su varianza  $\sigma^2$ . Entonces, por la desigualdad de Chebyshev se tiene que:

$$P(|S_n - \mu| \geq k) \leq \frac{\sigma^2}{nk^2}$$

$$P(|S_n - \mu| \geq k) \leq \delta \quad \text{si} \quad n \geq \frac{\sigma^2}{\delta k^2}$$

### Ejemplo

Supóngase que  $X_1, X_2, \dots, X_n$  son una muestra de variables aleatorias independientes con distribución de Bernoulli, de tal forma que toman el valor 1 con probabilidad  $p=0.5$ .

¿Cuál debe ser el tamaño de la muestra para poder garantizar que la probabilidad de que la diferencia entre la media aritmética  $S_n$  y su valor esperado (que exceda en más de 0,1), sea menor o igual que 0,01?

### Solución

Tenemos que  $E(X)=\mu=p=0,5$  y que  $\text{Var}(X)=\sigma^2=p(1-p)=0,25$ . Por la desigualdad de Chebyshev, para cualquier  $k>0$  tenemos que:

$$P(|S_n - \mu| \geq k) \leq \frac{p(1-p)}{nk^2}$$

Ahora, tomando  $k=0,1$  y  $\delta=0,01$ , se tiene que:

$$P(|S_n - 0.5| \geq 0.1) \leq 0.01 \quad \text{cuando} \quad n \geq \frac{0.5^2}{0.01^2} = 2500$$

De esta manera se concluye que se necesita un tamaño de muestra de al menos 2500 para garantizar que la probabilidad del evento  $|S_n - 0,5| \geq 0,1$  sea menor que 0,01.

## Desigualdades tipo Chebyshov

Existen diversas desigualdades relacionadas con la desigualdad de Chebyshov. Una de las más conocidas es la desigualdad de Markov:

$$P(X \geq k) \leq \frac{E(X^r)}{k^r}$$

En esta expresión X es una variable aleatoria no negativa con  $k, r > 0$ .

La desigualdad de Markov puede tomar distintas formas. Por ejemplo, sea Y una variable aleatoria no negativa (por lo que  $P(Y \geq 0) = 1$ ) y supongamos que  $E(Y) = \mu$  existe. Supongamos también que  $(E(Y))^r = \mu_r$  existe para algún entero  $r > 1$ . Entonces:

$$P(Y \geq A\mu) \leq \frac{1}{A} \quad \text{con } A > 0$$

$$P(Y \geq a) \leq \frac{\mu}{a} \quad \text{con } a > 0$$

$$P(Y \geq A\mu^r) \leq \frac{1}{A^r} \quad \text{con } A > 0$$

$$P(Y \geq a) \leq \frac{\mu_r}{a^r} \quad \text{con } a > 0$$

$$P(Y - \mu \geq h) \leq \frac{\mu}{\mu + h} \quad \text{con } h > 0$$

- 1.-Kai Lai Chung. Elementary Probability Theory with Stochastic Processes. Springer-Verlag New York Inc
- 2.-Kenneth.H. Rosen .Matemáticas Discretas y sus Aplicaciones . S.A.MCGRAW- HILL/INTERAMERICANA DE ESPAÑA.
- 3.-Paul L. Meyer. Probabilidad y Aplicaciones Estadísticas. S.A. ALHAMBRA MEXICANA.
- 4.-Seymour Lipschutz Ph.D. 2000 Problemas Resueltos de Matemática Discretas. McGRAW-HILL.

### 1.7 REGLA EMPIRICA

La regla empírica, a la que también se le conoce como la regla 68,5-95-99,7, constituye una manera útil de analizar datos estadísticos. Sin embargo, solo funciona para una distribución normal (la campana de Gauss) y solo es posible producir estimaciones. Será necesario que conozcas la media y la desviación estándar de los datos, pero, en caso de que vayas a emplear la regla empírica para una clase o un examen, se te deberá brindar esta información. Luego podrás usar esta regla para fines como estimar cuántos de los datos se encuentran dentro de un rango determinado.

**Traza y divide una distribución normal.** Haz un bosquejo de una curva normal cuyo punto más alto se encuentre en el centro y cuyos extremos se inclinen hacia abajo y vayan estrechándose de manera simétrica a la izquierda y la derecha. Luego traza varias líneas verticales que se intersequen con la curva de la siguiente manera:<sup>[1]</sup>

- Una línea debe dividir la curva por la mitad.
- Traza 3 líneas a la derecha de esta línea central y otras 3 a la izquierda. Estas líneas deben dividir cada una de las mitades de la curva en tres secciones espaciadas de manera uniforme y una sección pequeña en el extremo.

**Escribe los valores de la distribución normal en las líneas divisorias.** Marca la línea que esté en el centro con la media de tus datos y luego suma las desviaciones estándar para obtener los valores correspondientes a las 3 líneas a la derecha. Resta las desviaciones estándar a la media y obtendrás los valores correspondientes a las 3 líneas a la izquierda. Por ejemplo:<sup>[2]</sup>

- Imagina que la media de los datos es 16 y la desviación estándar es 2. Marca la línea central con el número 16.

- Suma las desviaciones estándar para marcar la primera línea a la derecha del centro con el número 18, la siguiente de la derecha con el número 20 y la línea en el extremo derecho con el número 22.
- Resta las desviaciones estándar para marcar la primera línea a la izquierda del centro con el número 14, la siguiente de la izquierda con el número 12 y la línea en el extremo izquierdo con el número 10.

**Marca los porcentajes de cada sección.** La regla empírica del punto base es fácil de comprender: el 68 % de los puntos de datos para una distribución normal se encontrarán dentro de una desviación estándar de la media, el 95 % dentro de dos desviaciones estándar y el 99,7 % dentro de tres desviaciones estándar. Puedes recordártelo marcando cada sección con un porcentaje de la siguiente forma:<sup>[3]</sup>

- Cada sección que se encuentre inmediatamente a la derecha y a la izquierda de la línea central contendrá el 34 %, sumando un total de 68.
- Las secciones siguientes a la derecha y a la izquierda contendrán cada una el 13,5 %. Súmalas al 68 % y obtendrás el 95 % de los datos.
- Las secciones siguientes a cada lado contendrán cada una 2,35% de los datos. Súmalos al 95 % y obtendrás el 99,7 % de los datos.
- Las puntas diminutas restantes de los datos a la izquierda y a la derecha contienen cada una 0,15 % de los datos restantes, lo cual suma un total de 100 %.

**Encuentra las distribuciones de tus datos.** Utiliza la media y la regla empírica para encontrar las distribuciones de los datos a 1, 2 y 3 desviaciones estándar de la media. Toma nota de ellos en la curva como referencia. Por ejemplo, imagina que vas a analizar el peso de una población de gatos, siendo la media del peso 4 kg y la desviación estándar 0,5 kg. Ten en cuenta lo siguiente:<sup>[4]</sup>

- 1 desviación estándar por encima de la media sería 4,5 kg y una desviación estándar por debajo de la media sería 3,5 kg.
- 2 desviaciones estándar por encima de la media sería 5 kg y 2 desviaciones estándar por debajo de la media sería 3 kg.
- 3 desviaciones estándar por encima de la media sería 5,5 kg y 3 desviaciones estándar por debajo de la media sería 2,5 kg.

**Determina la sección de la curva que la pregunta te pida que analices.** Cuando hayas dispuesto la curva, puedes resolver preguntas de análisis de datos mediante la regla empírica y la aritmética simple. Para empezar, lee con cuidado la pregunta para así determinar cuáles son las secciones con las que debes trabajar. Porejemplo.<sup>[5]</sup>

- Imagina que se te pide que encuentres el peso superior y el inferior para el 68 % de una población de gatos. Será necesario que trabajes con las dos secciones más cercanas al centro, en donde se encontrará el 68 % de los datos.
- De forma similar, imagina que la media del peso es 4 kg y la desviación estándar es 0,5 kg. En caso de que se te pida encontrar la proporción de gatos cuyo peso supere los 5 kg, deberás trabajar con la sección en el extremo derecho (a 2 desviaciones estándar de la media).

**Encuentra el porcentaje de los datos que estén dentro de un rango determinado.** En caso de que se te pida encontrar el porcentaje de la población dentro de un rango determinado, lo único que debes hacer es sumar los porcentajes dentro de un conjunto dado de desviaciones estándar. Por ejemplo, si lo que se te pide es encontrar el porcentaje de gatos cuyo peso sea de entre 3,5 y 5 kg, siendo la media de 4 kg y la desviación estándar de 0,5 kg, deberás tener en cuenta lo siguiente:<sup>[6]</sup>

- 2 desviaciones estándar por encima de la media equivaldrán a 5 kg, mientras que 1 desviación estándar por debajo de la media equivaldrá a 3,5 kg.
- Esto quiere decir el 81,5 % (68 % + 13,5 %) de los gatos pesa entre 3,5 y 5 kg.

**Encuentra puntos y rangos de datos empleando los porcentajes de las secciones.** Utiliza la información que te brinden las distribuciones de porcentajes y las desviaciones estándar para encontrar los límites superiores e inferiores para secciones de tus datos. Por ejemplo, una pregunta acerca de tus datos sobre el peso de los gatos podría ser —¿Cuál es el límite superior de peso del 2,5 % más bajo de los gatos? .<sup>[7]</sup>

- El 2,5 % más bajo de los datos se encontraría 2 desviaciones estándar por debajo de la media.
- Si es que la media es 4 kg y la desviación estándar es 0,5 kg, el peso del 2,5 % más bajo de los gatos será 3 kg o menos ( $4 - 0,5 \times 2$ ).



## **BIBLIOGRAFIA:**

<https://es.wikihow.com/usar-la-regla-emp%C3%A9rica>

1. ↑ <http://www.stat119review.com/more-material/normal-distribution/empirical-rule/solving-empirical-rule-questions>
2. ↑ <https://www.khanacademy.org/math/probability/normal-distributions-a2/normal-distributions-a2ii/v/ck12-org-normal-distribution-problems-empirical-rule>
3. ↑ <http://www.stat119review.com/more-material/normal-distribution/empirical-rule/solving-empirical-rule-questions>
4. ↑ <https://www.youtube.com/watch?v=T7-eeg6rhjY>
5. ↑ <http://www.stat119review.com/more-material/normal-distribution/empirical-rule/solving-empirical-rule-questions>
6. ↑ <https://www.khanacademy.org/math/probability/normal-distributions-a2/normal-distributions-a2ii/v/ck12-org-normal-distribution-problems-empirical-rule>
7. ↑ [https://www.nku.edu/~statistics/212\\_Using\\_the\\_Empirical\\_Rule.htm](https://www.nku.edu/~statistics/212_Using_the_Empirical_Rule.htm)

## UNIDAD II

### 2.1 TEORIA DE LA PROBABILIDAD

El concepto de probabilidad nace con el deseo del hombre de conocer con certeza los eventos futuros. Es por ello que el estudio de probabilidades surge como una herramienta utilizada por los nobles para ganar en los juegos y pasatiempos de la época. El desarrollo de estas herramientas fue asignado a los matemáticos de la corte.

Con el tiempo estas técnicas matemáticas se perfeccionaron y encontraron otros usos muy diferentes para la que fueron creadas. Actualmente se continúa con el estudio de nuevas metodologías que permitan maximizar el uso de la computación en el estudio de las probabilidades disminuyendo, de este modo, los márgenes de error en los cálculos.

A través de la historia se han desarrollado tres enfoques conceptuales diferentes para definir la probabilidad y determinar los valores de probabilidad:

#### El enfoque clásico

Dice que si hay  $x$  posibles resultados favorables a la ocurrencia de un evento  $A$  y  $z$  posibles resultados desfavorables a la ocurrencia de  $A$ , y todos los resultados son igualmente posibles y mutuamente excluyente (no pueden ocurrir los dos al mismo tiempo), entonces la probabilidad de que ocurra  $A$  es:

$$P(A) = \frac{x}{(x+z)}$$

El enfoque clásico de la probabilidad se basa en la suposición de que cada resultado sea igualmente posible.

Este enfoque es llamado enfoque a priori porque permite, (en caso de que pueda aplicarse) calcular el valor de probabilidad antes de observar cualquier evento de muestra.

### **Ejemplo:**

Si tenemos en una caja 15 piedras verdes y 9 piedras rojas. La probabilidad de sacar una piedra roja en un intento es:

$$P(A) = \frac{9}{9+15} = 0.375 \text{ o } 37.5\%$$

### **El enfoque de frecuencia relativa**

También llamado Enfoque Empírico, determina la probabilidad sobre la base de la proporción de veces que ocurre un evento favorable en un número de observaciones. En este enfoque no se utiliza la suposición previa de aleatoriedad. Porque la determinación de los valores de probabilidad se basa en la observación y recopilación de datos.

### **Ejemplo:**

Se ha observado que 9 de cada 50 vehículos que pasan por una esquina no tienen cinturón de seguridad. Si un vigilante de tránsito se para en esa misma esquina un día cualquiera

¿Cuál será la probabilidad de que detenga un vehículo sin cinturón de seguridad?

$$P(A) = \frac{9}{50} = 0.18 \text{ o } 18\%$$

Tanto el enfoque clásico como el enfoque empírico conducen a valores objetivos de probabilidad, en el sentido de que los valores de probabilidad indican a largo plazo la tasa relativa de ocurrencia del evento.

### **El enfoque subjetivo**

Dice que la probabilidad de ocurrencia de un evento es el grado de creencia por parte de un individuo de que un evento ocurra, basado en toda la evidencia a su disposición. Bajo esta premisa se puede decir que este enfoque es adecuado cuando solo hay una oportunidad de ocurrencia del evento. Es decir, que el evento ocurrirá o no ocurrirá esa sola vez. El valor de probabilidad bajo este enfoque es un juicio personal.

### **Concepto de Probabilidad**

Se define como cálculo de probabilidad al conjunto de reglas que permiten determinar si un fenómeno ha de producirse, fundando la suposición en el cálculo, las estadísticas o la teoría.

El objetivo de esta práctica es realizar varios experimentos de probabilidad, anotar los resultados y posteriormente compararlos con los resultados teóricos.

## Objetivos de las Probabilidades

El objetivo fundamental de la probabilidad, es la de mostrar al alumno la importancia y utilidad del Método Estadístico en el ámbito económico-empresarial. Con tal fin, el alumno deberá aprender a manejar los métodos y técnicas más adecuadas para el correcto tratamiento y análisis de la información proporcionada por los datos que genera la actividad económica.

Para ello se comienza afianzando los conocimientos que el alumno ya posee de Estadística Descriptiva, además de algunos conceptos nuevos relacionados con este tema.

## El valor de la probabilidad

El valor más pequeño que puede tener la probabilidad de ocurrencia de un evento es igual a 0, el cual indica que el evento es imposible, y el valor mayor es 1, que indica que el evento ciertamente ocurrirá. Entonces si decimos que  $P(A)$  es la probabilidad de ocurrencia de un evento  $A$  y  $P(A')$  la probabilidad de no-ocurrencia de  $A$ , tenemos que:

$$0 \leq P(A) \leq 1$$
$$P(A) + P(A') = 1$$

## Eventos mutuamente excluyentes y eventos no excluyentes

Dos o más eventos son mutuamente excluyentes o disjuntos, si no pueden ocurrir simultáneamente. Es decir, la ocurrencia de un evento impide automáticamente la ocurrencia del otro evento (o eventos).

### Ejemplo:

Allanzar una moneda solo puede ocurrir que salga cara o sello pero no los dos a la vez, esto quiere decir que estos eventos son excluyentes.

Dos o más eventos son no excluyentes, o conjuntos, cuando es posible que ocurran ambos. Esto no indica que necesariamente deban ocurrir estos eventos en forma simultánea.

### Ejemplo:

Si consideramos en un juego de domino sacar al menos un blanco y un seis, estos eventos son no excluyentes porque puede ocurrir que salga el seis blanco.

## Reglas de la Adición

La Regla de la Adición expresa que: la probabilidad de ocurrencia de al menos dos sucesos  $A$  y  $B$  es igual a:

$P(A \cup B) = P(A) \cup P(B) = P(A) + P(B)$  si A y B son mutuamente excluyentes

$P(A \cup B) = P(A) + P(B) - P(A \cap B)$  si A y B son no excluyentes

Siendo:  $P(A)$  = probabilidad de ocurrencia del evento A  $P(B)$  =

probabilidad de ocurrencia del evento B

$P(A \cap B)$  = probabilidad de ocurrencia simultánea de los eventos A y B

### Eventos Independientes

Dos o más eventos son independientes cuando la ocurrencia o no-ocurrencia de un evento no tiene efecto sobre la probabilidad de ocurrencia del otro evento (o eventos). Un caso típico de eventos independiente es el muestreo con reposición, es decir, una vez tomada la muestra se regresa de nuevo a la población donde se obtuvo.

#### Ejemplo:

lanzar al aire dos veces una moneda son eventos independientes por que el resultado del primer evento no afecta sobre las probabilidades efectivas de que ocurra cara o sello, en el segundo lanzamiento.

### Eventos dependientes

Dos o más eventos serán dependientes cuando la ocurrencia o no-ocurrencia de uno de ellos afecta la probabilidad de ocurrencia del otro (o otros). Cuando tenemos este caso, empleamos entonces, el concepto de probabilidad condicional para denominar la probabilidad del evento relacionado. La expresión  $P(A|B)$  indica la probabilidad de ocurrencia del evento A sí el evento B ya ocurrió.

**Se debe tener claro que  $A|B$  no es una fracción.**

$$P(A|B) = P(A \cap B)/P(B) \text{ o } P(B|A) = P(A \cap B)/P(A)$$

### Reglas de Multiplicación

Se relacionan con la determinación de la ocurrencia de conjunta de dos o más eventos. Es decir la intersección entre los conjuntos de los posibles valores de A y los valores de B, esto quiere decir que la probabilidad de que ocurran conjuntamente los eventos A y B es:

$$P(A \cap B) = P(A \cap B) = P(A)P(B) \text{ si A y B son independientes}$$

$P(A \cap B) = P(A)P(B|A)$  si A y B son dependientes  $P(A \cap B) = P(A)$

$P(A \cap B) = P(B)P(A|B)$  si A y B son dependientes **Distribución de**

**probabilidad normal**

Es una distribución de probabilidad continua que es tanto simétrica como mesocurtica. La curva que representa la distribución de probabilidad normal se describe generalmente como en forma de campana. Esta distribución es importante en inferencia estadística por tres razones diferentes:

1. Se sabe que las medidas producidas en muchos procesos aleatorios siguen esta distribución.
2. Las probabilidades normales pueden utilizarse generalmente para aproximar otras distribuciones de probabilidad, tales como las distribuciones binomial y de Poisson.
3. Las distribuciones estadísticas tales como la media de la muestra y la proporción de la muestra, siguen a menudo la distribución normal, sin tener en cuenta la distribución de la población

Los valores de los parámetros de la distribución de probabilidad normal son  $\mu = 0$  y  $\sigma = 1$ . Cualquier conjunto de valores X normalmente distribuido pueden convertirse en valores normales estándar z por medio de la formula:

$$Z = \frac{X - \mu}{\sigma}$$

Haciendo posible el uso de la tabla de proporciones de área y hace innecesario el uso de la ecuación de la función de densidad de cualquier distribución normal dada.

Para aproximar las distribuciones discretas binomial y de Poisson se debe hacer:

Binomial	$\mu = np$	$\sigma = \sqrt{np(1-p)}$	Si $n > 30$ $.np > 5$ $n(1-p) > 5$
Poisson	$\mu = \lambda$	$\sigma = \sqrt{\lambda}$	

## Distribución de probabilidad exponencial

Si en el contexto de un proceso de Poisson ocurren eventos o éxitos en un espectro continuo de tiempo y espacio. Entonces la longitud del espacio o tiempo entre eventos sucesivos sigue una distribución de probabilidad exponencial. Puesto que el tiempo y el espacio son un espectro continuo, esta es una distribución continua.

Encaso de este tipo de distribución no vale la pena preguntarse ¿cuál es la probabilidad de que el primer pedido de servicio se haga exactamente de aquí a un minuto?. Mas bien debemos asignar un intervalo dentro del cual el evento puede ocurrir, preguntándonos, ¿cuál es la probabilidad de que el primer pedido se produzca en el próximo minuto?.

Dado que el proceso de Poisson es estacionario, la distribución exponencial se aplica ya sea cuando estamos interesados en el tiempo (o espacio) hasta el primer evento, el tiempo entre dos eventos sucesivos, o el tiempo hasta que ocurra el primer evento después de cualquier punto aleatoriamente seleccionado.

### Ejemplo:

Un departamento de mantenimiento recibe un promedio de 5 llamadas por hora. Comenzando en un momento aleatoriamente seleccionado, la probabilidad de que una llamada llegue dentro de media hora es:

Promedio 5 por hora, como el intervalo es media hora tenemos que  $\lambda = 2,5/\text{media hora}$ .  $P(T < 30 \text{ min.}) = 1 - e^{-5} = 1 - 0,08208 = 0,91792$

### BIBLIOGRAFIA:

- 1.-ÉREZ, C. (1995) *Análisis Estadístico con Statgraphics. Técnicas Básicas*. Madrid: Ra-Ma.
- 2.-TANUR, J. (1992) *La Estadística, una Guía de lo Desconocido*. Madrid: Alianza Editorial.

### **2.1.1 ENFOQUES DE PROBABILIDAD**

La probabilidad es la posibilidad que existe entre varias posibilidades, que un hecho o condición se produzcan. La probabilidad, entonces, mide la frecuencia con la cual se obtiene un resultado en oportunidad de la realización de un experimento sobre el cual se conocen todos los resultados posibles gracias a las condiciones de estabilidad que el contexto supone de antemano.

Expresado matemáticamente, es igual al número de formas que un evento específico puede ocurrir, dividido por el número total de posibles eventos. Por ejemplo, si tienes una bolsa con tres canicas, una azul y dos verdes, la probabilidad de tomar una canica azul sin mirar es de  $1/3$ . Hay sólo un resultado posible de que se seleccione la canica azul, pero hay tres posibles resultados en total, azul, verde, verde. Usando el mismo razonamiento, la probabilidad de tomar una canica verde es de  $2/3$ .

La gran aliada de la probabilidad es la llamada teoría de la probabilidad, ya que gracias a lo que esta postula y sostiene, es que los seres humanos podemos anticiparnos a que algunos sucesos potenciales ocurran finalmente. La mencionada teoría es muy utilizada y consultada por disciplinas como ser la estadística, la filosofía, las matemáticas y la ciencia, para sacar conclusiones respecto de los sucesos potenciales que las ocupan.

La teoría de la probabilidad es un modelo matemático que se ocupa de analizar los fenómenos aleatorios; esto implica la contraposición respecto de los fenómenos ya determinados, que son aquellos en los cuales el resultado del experimento que se realiza, atendiendo a determinadas condiciones, produce un resultado único y previsible, que se repetirá la cantidad de veces que éste vuelva a hacerse, siempre y cuando se respeten las mismas condiciones.

Entonces, dentro de la teoría de la probabilidad se intenta determinar la cantidad de veces que puede un determinado resultado acontecer, con el fin de conocer que suceso es el más probable.



Por ejemplo, el agua que se calienta a 100 grados Celsius a nivel del mar se convierte en vapor: éste es un fenómeno ya determinado. En tanto, los aleatorios, que son de los que se ocupa la teoría de la probabilidad, podrán realizarse miles de veces bajo las mismas circunstancias pero siempre tendrán como resultado un variado conjunto de alternativas. Un clarísimo ejemplo resulta ser las diversas posibilidades y combinaciones que permite el lanzamiento de dados cuando se está jugando a la generala.

La probabilidad está absolutamente inmersa en nuestro día a día como parte integrante de una sociedad y comunidad determinada, ya que en el análisis de riesgos y en el comercio de materias primas, la probabilidad, tiene una incidencia y una importancia vital.

Existen tres tipos de enfoques de Probabilidad:

- Clásico
- Relativo
- Subjetivo

#### **Clásico:**

Los resultados de un experimento son igualmente viables, es decir, tienen teóricamente las mismas posibilidades de ocurrir.

En este caso la probabilidad de ocurrencia de un evento será:

Número de resultados en los que se presenta el evento / número total de resultados posibles

Por ejemplo, la probabilidad de que en una baraja francesa de 52 cartas salga el cinco de trébol es de  $1/52$ .

### **De Frecuencia Relativa:**

La probabilidad de que un evento suceda se determina observando eventos similares en el pasado. Este método utiliza la frecuencia relativa de las presentaciones pasadas de un evento como una probabilidad. Determinamos qué tan frecuente ha sucedido algo en el pasado y usamos esa cifra para predecir la probabilidad de que suceda de nuevo en el futuro.

En este caso la probabilidad de ocurrencia de un evento será:

Número de resultados esperados ocurridos en el pasado / número total de experimentos adelantados

Por ejemplo, la probabilidad de que Brasil gané el mundial de Rusia 2018 es de 5 mundiales ganados anteriormente / 20 mundiales que se han celebrado en total.

### **De Frecuencia Subjetiva:**

Se puede definir como la probabilidad asignada a un evento por parte de un individuo, basada en la evidencia que se tenga disponible. Esa evidencia puede presentarse en forma de frecuencia relativa de presentación de eventos pasados o puede tratarse simplemente de una creencia meditada.

### **BIBLIOGRAFIA:**

1.-URIEL, E. y MUÑIZ, M. (1988) *Estadística Económica y Empresarial. Teoría y ejercicios*. Madrid: AC.

2.-URIEL, E. y PEIRÓ, A. (2000) *Introducción al análisis de series temporales*. Madrid: AC. 3.-[http://soy-  
staff.blogspot.com/2015/10/aspectos-generales-de-la-probabilidad.html](http://soy-<br/>staff.blogspot.com/2015/10/aspectos-generales-de-la-probabilidad.html)

## 2.1.2 ESPACIO MUESTRAL

En cualquier experimento aleatorio la primera cosa que nos preguntamos es sobre lo que puede pasar. ¿Qué resultados puede ofrecer y cuáles no? Sería muy interesante disponer de todo el abanico de posibles resultados. En este sentido, al conjunto formado por todos los posibles resultados elementales de un experimento aleatorio se le denomina espacio muestral de dicho experimento. Dependiendo de como sea este conjunto, los espacios muestrales pueden ser:

- Espacio muestral discreto finito. Consta de un número finito de elementos, por ejemplo lanzar un dado.
- Espacio muestral discreto infinito. Consta de un número infinito numerable de elementos, por ejemplo lanzar un dado hasta que salga un cinco.
- Espacio muestral continuo. Consta de un número infinito no numerable de elementos, por ejemplo todas las medidas posibles de espárragos extraídos aleatoriamente de una población.

Consideremos por ejemplo:

1. El experimento consistente en el lanzamiento de un dado y anotar el resultado de la cara superior. El espacio muestral sería:

$$E = \{ 1, 2, 3, 4, 5, 6 \}$$

2. El experimento consistente en el lanzamiento de dos monedas al aire. El espacio muestral o conjunto de todos los resultados elementales posibles sería:

$$E = \{ CCC, CCF, CFC, FCC, CFF, FCF, FFC, FFF \}$$

3. El experimento consistente en elegir aleatoriamente cualquier número de tres cifras mediante la extracción con reemplazamiento de bolas de una urna en la que aparecen las diez cifras significativas. En este caso el espacio muestral sería:

$$E = \{ 000, 001, \dots, 999 \}$$

4. El experimento consistente en el lanzamiento de dos dados de los que después se escogera la mejor de las puntuaciones. El espacio muestral sería:

$$E = [ 1, 2, 3, 4, 5, 6 ]$$

5. El experimento consistente en abrir aleatoriamente un libro y anotar después la primera letra de la página de la izquierda. El espacio muestral en estos casos sería:

$$E = \{ A, B, \dots, Z \}$$

Los ejemplos que podrían exponerse son innumerables y seguro que ya estás pensando en diversas situaciones. No obstante, de partida, queremos que te fijas y pienses en lo que te vamos a exponer. Observa el ejemplo (1) y el (4), el espacio muestral es el mismo, pero ¿puede considerarse el mismo?, esto es, los sucesos que aparecen sí son los mismos pero la ocurrencia de cada suceso en el experimento (1) no tiene el mismo comportamiento que la ocurrencia de cada suceso en el experimento (4)

#### **BIBLIOGRAFIA:**

[https://proyectodescartes.org/iCartesiLibri/materiales\\_didacticos/EstadisticaProbabilidadInferencia/Probabilidad/2\\_1ExperimentosAleatorios/index.html](https://proyectodescartes.org/iCartesiLibri/materiales_didacticos/EstadisticaProbabilidadInferencia/Probabilidad/2_1ExperimentosAleatorios/index.html)

### **2.1.3 EXPERIMENTOS SIMPLES Y COMPLEJOS**

Primero necesitamos introducir algunos términos. Cuando trabajamos con probabilidad, una acción aleatoria o serie de acciones se llama experimento. Un resultado es la consecuencia de un experimento, y un evento es una colección particular de resultados. Los eventos usualmente son descritos usando una característica común de los resultados.

Apliquemos este lenguaje para ver cómo funcionan los términos en la práctica. Algunos juegos requieren lanzar un dado de seis lados, numerado del 1 al 6. La tabla siguiente ilustra el uso de experimento,

resultado, y evento en ese juego:

Experimento	Resultados	Eventos
Lanzar un dado	Existen 6 resultados posibles: {1, 2, 3, 4, 5, 6}	Sacar un número par: {2, 4, 6} Sacar un 3: {3} Sacar un 1 o un 3: {1, 3} Sacar un 1 y un 3: {} (Sólo puede salir un número, por lo que esto es imposible. El evento no contiene resultados.)

Nota que una colección de resultados se pone en corchetes y separado por comas.

Un evento simple es un evento con *un* solo resultado. Sacar un 1 sería un evento simple, porque existe sólo un resultado que funciona: 1. Sacar más que 5 también sería un evento simple, porque el evento incluye sólo al 6 como un resultado válido.

#### 2.1.4 LEYES DE PROBABILIDAD

La probabilidad es un método por el cual se obtiene la frecuencia de un suceso determinado mediante la realización de un experimento aleatorio, del que se conocen todos los resultados posibles, bajo condiciones *suficientemente* estables. La teoría de la probabilidad se usa extensamente en áreas como la estadística, la física, la matemáticas, las ciencias y la filosofía para sacar conclusiones sobre la probabilidad discreta de sucesos potenciales y la mecánica subyacente discreta de sistemas complejos.

La probabilidad constituye un importante parámetro en la determinación de las diversas casualidades obtenidas tras una serie de eventos esperados dentro de un rango estadístico.

Existen diversas formas como método abstracto, como la teoría de Dempster y la teoría de la relatividad numérica, esta última con un alto grado de aceptación si se toma en cuenta que disminuye considerablemente las posibilidades hasta un nivel mínimo ya que somete a todas las antiguas reglas a una simple ley de relatividad.

La probabilidad de un evento se denota con la letra  $p$  y se expresa en términos de una fracción y no en porcentajes, por lo que el valor de  $p$  cae entre 0 y 1. Por otra parte, la probabilidad de que un evento —no ocurra— equivale a 1 menos el valor de  $p$  y se denota con la letra  $q$

$$P(Q) = 1 - P(E)$$

Los tres métodos para calcular las probabilidades son la regla de la adición, la regla de la multiplicación.

#### Regla de la adición

La regla de la adición o regla de la suma establece que la probabilidad de ocurrencia de cualquier evento en particular es igual a la suma de las probabilidades individuales, si es que los eventos son mutuamente excluyentes, es decir, que dos no pueden ocurrir al mismo tiempo.

$P(A \cup B) = P(A) \cup P(B) = P(A) + P(B)$  si A y B son mutuamente excluyentes.  $P(A \cup B) = P(A) + P(B) - P(A \cap B)$  si A y B son no excluyentes.

Siendo:  $P(A)$  = probabilidad de ocurrencia del evento A.  $P(B)$  = probabilidad de ocurrencia del evento B.  $P(A \cap B)$  = probabilidad de ocurrencia simultánea de los eventos A y B.

#### Regla de la multiplicación

La regla de la multiplicación establece que la probabilidad de ocurrencia de dos o más eventos estadísticamente independientes es igual al producto de sus probabilidades individuales.

$P(A \cap B) = P(A)P(B)$  si A y B son independientes.  $P(A \cap B) = P(A)P(B|A)$  si A y B son dependientes

La regla de Laplace establece que:

- La probabilidad de ocurrencia de un suceso *imposible* es 0.
- La probabilidad de ocurrencia de un suceso *seguro* es 1, es decir,  $P(A) = 1$ .

Para aplicar la regla de Laplace es necesario que los experimentos den lugar a sucesos equiprobables, es decir, que todos tengan o posean la misma probabilidad.

- La probabilidad de que ocurra un suceso se calcula así:

$$P(A) = \frac{\text{Nº de casos favorables}}{\text{Nº de resultados posibles}}$$

Esto significa que: la probabilidad del evento A es igual al cociente del número de casos favorables (los casos donde sucede A) sobre el total de casos posibles.-

-Aplicacion:

Dos aplicaciones principales de la teoría de la probabilidad en el día a día son en el análisis de riesgo y en el comercio de los mercados de materias. Los gobiernos normalmente aplican métodos probabilísticos en regulacion ambiental donde se les llama —análisis de vias de dispersion , y a menudo miden el bienestar usando métodos que son estocásticos por naturaleza, y escogen qué proyectos emprender basándose en análisis estadísticos de su probable efecto en la población como un conjunto. No es correcto decir que la estadística está incluida en el propio modelado, ya que típicamente los análisis de riesgo son para una única vez y por lo tanto requieren más modelos de probabilidad fundamentales, por ej. —la probabilidad de otro 11-S . Una ley de numeros pequeños tiende a aplicarse a todas aquellas elecciones y percepciones del efecto de estas elecciones, lo que hace de las medidas probabilísticas un tema político.

Un buen ejemplo es el efecto de la probabilidad percibida de cualquier conflicto generalizado sobre los precios del petróleo en Oriente Medio – que producen un efecto

dominó en la economía en conjunto. Un cálculo por un mercado de materias primas en que la guerra es más probable en contra de menos probable probablemente envía los precios hacia arriba o hacia abajo e indica a otros comerciantes esa opinión. Por consiguiente, las probabilidades no se calculan independientemente y tampoco son necesariamente muy racionales. La teoría de las fianzas conductuales surgió para describir el efecto de este pensamiento de grupo en el precio, en la política, y en la paz y en los conflictos.

Se puede decir razonablemente que el descubrimiento de métodos rigurosos para calcular y combinar los cálculos de probabilidad ha tenido un profundo efecto en la sociedad moderna. Por consiguiente, puede ser de alguna importancia para la mayoría de los ciudadanos entender cómo se calculan los pronósticos y las probabilidades, y cómo contribuyen a la reputación y a las decisiones, especialmente en una democracia.

Otra aplicación significativa de la teoría de la probabilidad en el día a día es en la fiabilidad. Muchos bienes de consumo, como los automoviles y la electrónica de consumo, utilizan la teoría de la fiabilidad en el diseño del producto para reducir la probabilidad de avería. La probabilidad de avería también está estrechamente relacionada con la garantía del producto.

Se puede decir que no existe una cosa llamada probabilidad. También se puede decir que la probabilidad es la medida de nuestro grado de incertidumbre, o esto es, el grado de nuestra ignorancia dada una situación. Por consiguiente, puede haber una probabilidad de 1 entre 52 de que la primera carta en un baraja sea la J de diamantes. Sin embargo, si uno mira la primera carta y la reemplaza, entonces la probabilidad es o bien 100% ó 0%, y la elección correcta puede ser hecha con precisión por el que ve la carta. La física moderna proporciona ejemplos importantes de situaciones determinísticas donde sólo la descripción probabilística es factible debido a información incompleta y la complejidad de un sistema así como ejemplos de fenómenos realmente aleatorios.



## **BIBLIOGRAFIA:**

1.-Anderson D., Sweeney D., Williams T. Estadística para la administración y economía. Décima edición. Cengage Learning. 2008

2.-Berenson M., Levine D., Krehbiel T. Estadística para administración. Segunda edición. Prentice Hall. 2000

### **2.1.5 TABLAS DE CONTINGENCIA**

Una tabla de contingencia es una tabla que cuenta las observaciones por múltiples variables categóricas. Las filas y columnas de las tablas corresponden a estas variables categóricas.

Por ejemplo, después de una elección reciente entre dos candidatos, una encuesta de salida registró el sexo y el voto de 100 electores seleccionados de manera aleatoria y los datos se tabularon de la siguiente manera:

	Candidato A	Candidato B	Todo
Hombre	28	20	48
Mujer	39	13	52
Todo	67	33	100

Esta tabla de contingencia cuenta las respuestas según sexo y voto. El conteo en la intersección de la fila  $i$  y la columna  $j$  se denota como  $n_{ij}$ , y representa el número de observaciones que muestra esa combinación de niveles. Por ejemplo,  $n_{1,2}$  muestra el número de encuestados masculinos que votaron por el Candidato B.

La tabla también incluye los totales marginales para cada nivel de las variables. Los totales marginales para las filas muestran que 52 de los encuestados fueron mujeres. Los totales marginales para las columnas muestran que 67 encuestados votaron por el Candidato A. Además, el total general muestra que el tamaño de la muestra es 100.

Las tablas de contingencia también pueden revelar asociaciones entre las dos variables. Utilice una prueba de chi-cuadrada o una prueba exacta de Fisher para determinar si los conteos observados difieren significativamente de los conteos esperados bajo la hipótesis nula de que no existe asociación. Por ejemplo, usted podría probar si existe una asociación entre sexo y voto.

Las tablas de contingencia más simples son tablas de dos factores que cuentan las respuestas según dos variables. Usted puede categorizar las observaciones según tres o más variables al "cruzarlas". En el ejemplo de votación anterior, las respuestas también podrían clasificarse según el estatus de empleo de la manera siguiente:

	Candidato A	Candidato B	Total
Hombre / empleado	18	19	37
Hombre / desempleado	10	1	11
Mujer / empleada	33	10	43
Mujer /desempleada	6	3	9
Total	67	33	100

Un análisis de correspondencia simple puede detectar asociaciones en las tablas de contingencia que categorizan los datos por más de dos variables. Para realizar un análisis de correspondencia simple en Minitab, elija Estadísticas>Análisis multivariado>Análisis de correspondencia simple.

Calcular la relación de probabilidades y el intervalo de confianza para una tabla de contingencia 2 X 2  
 Usted puede usar Estadísticas > Regresión > Regresión logística binaria > Ajustar modelo logístico binario para calcular la relación de probabilidades y el intervalo de confianza.

Por ejemplo, usted está investigando la relación entre el uso de aspirina y ataques al corazón, y desea calcular la relación de probabilidades y el intervalo de confianza de la relación de probabilidades para la siguiente tabla de contingencia 2 X 2:

	Ataque al corazón	Sin ataque al corazón
Placebo	189	10845
Aspirina	104	10933

1. Ingrese los siguientes datos en Minitab:

C1	C2	C3
Grupo	Ataque al corazón	Conteo
Placebo	Sí	189
Placebo	No	10845
Aspirina	Sí	104
Aspirina	No	10933

2. Elija Estadísticas > Regresión > Regresión logística binaria > Ajustar modelo logístico binario.

3. En Respuesta, ingrese C2 y en Frecuencia, ingrese C3.

4. En Predictores categóricos, ingrese C1. Haga clic en Aceptar.

Regresión logística binaria: Ataque al corazón vs. Grupo

## Relaciones de probabilidades para predictores categóricos

	Relación de	
NivelA	NivelB probabilidades	ICde95%
Grupo		
Placebo	Aspirina	1.8321 (1.4400,2.3308)

### Relación de probabilidades para nivel A relativo a nivel B

La relación de probabilidades es 1,8321. Esto significa que una persona que tome el placebo tiene una probabilidad 1,8321 veces mayor de sufrir un ataque al corazón que una persona que tome aspirina. Usted puede estar 95% seguro de que el valor real de la relación de probabilidades está entre 1,44 y 2,3308.

### **BIBLIOGRAFIA:**

1. Blair C., Taylor R. Bioestadística. Peason. Prentice Hall. 2008
2. Daniel W. Bioestadística. Cuarta edición. Limusa Wiley. 2006

## 2.1.6 TEOREMA DE BAYES

**El teorema de Bayes es utilizado para calcular la probabilidad de un suceso, teniendo información de antemano sobre ese suceso.**

Podemos calcular la probabilidad de un suceso A, sabiendo además que ese A cumple cierta característica que condiciona su probabilidad. El teorema de Bayes entiende la probabilidad de forma inversa al teorema de la probabilidad total. El teorema de la probabilidad total hace inferencia sobre un suceso B, a partir de los resultados de los sucesos A. Por su parte, Bayes calcula la probabilidad de A condicionado a B.

El teorema de Bayes ha sido muy cuestionado. Lo cual se ha debido, principalmente, a su mala aplicación. Ya que, mientras se cumplan los supuestos de sucesos disjuntos y exhaustivos, el teorema es totalmente válido.

Fórmula del teorema de Bayes

Para calcular la probabilidad tal como la definió Bayes en este tipo de sucesos, necesitamos una fórmula. La fórmula se define matemáticamente como:

$$P[A_n/B] = \frac{P[B/A_n] \cdot P[A_n]}{\sum P[B/A_i] \cdot P[A_i]}$$

Donde B es el suceso sobre el que tenemos información previa y A(n) son los distintos sucesos condicionados. En la parte del numerador tenemos la probabilidad condicionada, y en la parte de abajo la probabilidad total. En cualquier caso, aunque la fórmula parezca un poco abstracta, es muy sencilla. Para demostrarlo, utilizaremos un ejemplo en el que en lugar de A(1), A(2) y A(3), utilizaremos directamente A, B y C.

Ejemplo del teorema de Bayes

*Una empresa tiene una fábrica en Estados Unidos que dispone de tres máquinas A, B y C, que producen envases para botellas de agua. Se sabe que la máquina A produce un 40% de la cantidad total, la máquina B un 30%, y la máquina C un 30%. También se sabe que cada máquina produce envases defectuosos. De tal manera que la máquina A produce un 2% de envases defectuosos sobre el total de su producción, la máquina B un 3%, y la máquina C un 5%. Dicho esto, se plantean dos cuestiones:*

$$P(A)=0,40 \quad P(D/A) = 0,02$$

$$P(B)=0,30 \quad P(D/B) = 0,03$$

$$P(C)=0,30 \quad P(D/C) = 0,05$$

1. Si un envase ha sido fabricado por la fábrica de esta empresa en Estados Unidos ¿Cuál es la probabilidad de que sea defectuoso?

Se calcula la probabilidad total. Ya que, a partir los diferentes sucesos, calculamos la probabilidad de que sea defectuoso.

$$P(D) = [P(A) \times P(D/A)] + [P(B) \times P(D/B)] + [P(C) \times P(D/C)] = [0,4 \times 0,02] + [0,3 \times 0,03] + [0,3 \times 0,05] = 0,032$$

Expresado en porcentaje, diríamos que la probabilidad de que un envase fabricado por la fábrica de esta empresa en Estados Unidos sea defectuoso es del 3,2%.

2. Siguiendo con la pregunta anterior, si se adquiere un envase y este es defectuoso ¿Cuáles es la probabilidad de que haya sido fabricado por la máquina A? ¿Y por la máquina B? ¿Y por la máquina C?

Aquí se utiliza el teorema de Bayes. Tenemos información previa, es decir, sabemos que el envase es defectuoso. Claro que, sabiendo que es defectuoso, queremos saber cual es la probabilidad de que se haya producido por una de las máquinas.

$$P(A/D) = [P(A) \times P(D/A)] / P(D) = [0,40 \times 0,02] / 0,032 = 0,25$$

$$P(B/D) = [P(B) \times P(D/B)] / P(D) = [0,30 \times 0,03] / 0,032 = 0,28$$

$$P(C/D) = [P(C) \times P(D/C)] / P(D) = [0,30 \times 0,05] / 0,032 = 0,47$$

Sabiendo que un envase es defectuoso, la probabilidad de que haya sido producido por la máquina A es del 25%, de que haya sido producido por la máquina B es del 28% y de que haya sido producido por la máquina C es del 47%.

**BIBLIOGRAFIA:**

1. Meyer P. Probabilidad y Aplicaciones estadísticas. Edición revisada. Addison Wesley Logman. 1998
2. Montgomery D., Diseño y análisis de experimentos. Segunda edición. Limusa Wiley. 2006