

Estadísticas para Negocios Internacionales



MAWIL

Publicaciones Impresas
y Digitales

ESTADÍSTICAS PARA NEGOCIOS INTERNACIONALES



ESTADÍSTICAS PARA NEGOCIOS INTERNACIONALES

AUTORES

Ec. Luis Agustin Fajardo Vaca Mg.

Magister en Administración de Pequeñas y Medianas Empresas;

Economista; Licenciado en Ciencias Políticas y Sociales

Docente Universidad de Guayaquil

luis.fajardova@ug.edu.ec

Ing. Johanna Estefania Rangel Saltos Mg.

Magister en Ciencias con mención en Economía y Gestión Empresarial;

Ingeniera en Gestión Empresarial Internacional

Docente Universidad de Guayaquil

johanna.rangels@ug.edu.ec

Ec. Nancy Rocio Castillo Castro Mg.

Magister en Finanzas y Proyectos Corporativos; Economista

Docente Universidad de Guayaquil

nancy.castilloc@ug.edu.ec

Ing. Karen Giselle Ochoa Quirola Mg.

Magister en Administración de Empresas con mención en

Recursos Humanos; Ingeniera en Gestión Empresarial

Docente Universidad de Guayaquil

karen.ochoaq@ug.edu.ec

ESTADÍSTICAS PARA NEGOCIOS INTERNACIONALES

REVISORES

Econ. Xavier Francisco Flores Torres Msc.
xavier.florest@ug.edu.ec

Docente de la Universidad de Guayaquil
Facultad de Ciencias Administrativas
Coordinador de Logística Nivelación y Admisión

Ing. Verónica Patricia Silva Ortega MGP.
veronica.silvaor@ug.edu.ec

Docente de la Universidad de Guayaquil
Doctorando Pontificia
Universidad Católica de Santa María de Buenos Aires
en Administración de Empresas

DATOS DE CATALOGACIÓN

AUTORES: Luis Agustín Fajardo Vaca
Johanna Estefanía Rangel Saltos
Nancy Roció Castillo Castro
Karen Giselle Ochoa Quirola

Título: Estadísticas para negocios internacionales

Descriptor: Matemáticas; Probabilidades; Conjuntos; Variancia; Hipótesis.

Edición: 1^{era}

ISBN: 978-9942-787-57-6

Editorial: Mawil Publicaciones de Ecuador, 2018

Área: Educación Superior

Formato: 148 x 210 mm.

Páginas: 204

DOI: [10.26820/mawil/978-9942-787-57-6](https://doi.org/10.26820/mawil/978-9942-787-57-6)

Texto para Docentes y Estudiantes Universitarios

El proyecto didáctico *Estadísticas para negocios internacionales*, es una obra colectiva creada por sus autores y publicada por *MAWIL*; publicación revisada por el equipo profesional y editorial siguiendo los lineamientos y estructuras establecidos por el departamento de publicaciones de *MAWIL* de New Jersey.

© Reservados todos los derechos. La reproducción parcial o total queda estrictamente prohibida, sin la autorización expresa de los autores, bajo sanciones establecidas en las leyes, por cualquier medio o procedimiento.

*Director General: Mg. Vanessa Quishpe Morocho

*Dirección Central MAWIL: Office 18 Center Avenue Caldwell; New Jersey # 07006

*Gerencia Editorial MAWIL-Ecuador: Lenin Suasnabas

*Editor de Arte y Diseño: Alejandro Plúa

El pueblo más feliz es el que tenga mejor educados a sus hijos, en la instrucción del pensamiento, y en la dirección de los sentimientos. Un pueblo instruido ama el trabajo y sabe sacar provecho de él. Un pueblo virtuoso vivirá más feliz y más rico que otro lleno de vicios, y se defenderá mejor de todo ataque.”

José Martí Pérez.

ÍNDICE

ESTADÍSTICAS PARA NEGOCIOS INTERNACIONALES



www.mawil.us

ÍNDICE	PÁGINAS
UNIDAD I	
DEFINICIONES ESTADÍSTICAS Y PRESENTACIÓN DE DATOS ESTADÍSTICOS.....	8
UNIDAD II	
INDICADORES DE TENDENCIA CENTRAL Y DE VARIABILIDAD.....	62
UNIDAD III	
NUMERO ÍNDICES Y ANÁLISIS DE REGRESIÓN SIMPLE	71
UNIDAD IV	
TEORÍA DE LAS PROBABILIDADES Y DISTRIBUCIÓN DE PROBABILIDADES NORMAL.....	108
UNIDAD V	
INTRODUCCIÓN A LA DISTRIBUCIÓN DE PROBABILIDADES DE VARIABLES CONTINUAS Y DISTRIBUCIONES MUESTRALES	124
UNIDAD VI	
TÉCNICAS DE MUESTREO Y ESTIMACIÓN DE PARÁMETROS POBLACIONALES	137
UNIDAD VII	
TÉCNICAS DE MUESTREO Y ESTIMACIÓN DE PARÁMETROS POBLACIONALES	151
UNIDAD VIII	
ANÁLISIS DE VARIANZA (ANOVA) DE K MEDIAS Y REGRESIÓN MÚLTIPLE.....	162

UNIDAD I

DEFINICIONES ESTADÍSTICAS Y PRESENTACIÓN DE DATOS ESTADÍSTICOS



www.mawil.us

1.1. Concepto y definiciones de la estadística

La estadística siempre se ha pensado que es una relación de datos numéricos presentada de forma ordenada y sistemática; que ha sido consecuencia del concepto popular existente sobre el término, extendiéndose cada vez concebido por la influencia de nuestro entorno, debido a que frecuentemente es casi imposible que cualquier medio de difusión; periódico, radio, televisión, etc, no lo aborden diariamente comunicando diferentes tipo de informaciones estadísticas referidos a accidentes de tráfico, índices de crecimiento de población, turismo, tendencias políticas sociales, de bienestar de salud, protección, prevención entre otras.



“ESTADISTICA” Es atribuido a la palabra “ESTADO”; ya que esencialmente la función de los gobiernos entre otras cuestiones es llevar los registros de población, nacimientos, cosechas, impuestos y toda la información que engloba el estado, por ello, tradicionalmente se ha definido a la Estadística como un instrumento de compilación, organización, presentación y análisis de datos generalmente numéricos; ya quedando conceptualizado globalmente como la parte de las Matemáticas encargada del estudio de una específica característica en una población; recogiendo los datos; organizándolos en tablas; representándolos gráficamente y además analizándolos para sacar conclusiones de dicha población. En un sentido más abierto es promulgable decir que la Estadística se utiliza como tecnología al servicio de las ciencias donde la variabilidad y la incertidumbre forman parte de su naturaleza, planteándose en ese sentido que

la Estadística es Ciencia siendo su objetivo esencial reunir toda información cuantitativa correspondiente a individuos, grupos, series de hechos, etc., para la deducción de ello, gracias al análisis de estos datos; significados precisos o previsiones para el futuro; por lo en general, es la Ciencia que trata de la recopilación, organización presentación, análisis e interpretación de datos numéricos con la finalidad de tomar decisiones efectivas y pertinentes.

Aunque otros autores coinciden con las definiciones de la Estadística semejantes a las anteriores igualmente algunos otros no comparten las semejanzas : Mayormente la definen como; ciencia que tiene por objeto el estudio cuantitativo de los colectivos, otros como expresión cuantitativa del conocimiento disponible de manera adecuada para el escrutinio y análisis, aunque la más aceptada, sin embargo, es la de Minguez, el cual define la Estadística como manera: “La ciencia que posee por objeto la aplicación de las leyes de la cantidad a los hechos sociales para así medir su intensidad, deducción de las leyes que los rigen haciendo su predicción próxima.

La estadística para los negocios proporciona las herramientas posibles y necesarias para recopilar, analizar; presentación e interpretación de datos, en el campo de los negocios y la economía, facilitándole a quienes toman decisiones mejor comprensión del entorno comercial y económico, logrando ser eficientes en el desarrollo de sus tareas. En este sentido el término Estadística es referido a datos numéricos, como: promedios, medianas, porcentajes y números índices que coadyuvan al entendimiento de una gran gama de negocios y situaciones económicas: Observable es que, el campo de la estadística es meramente más que datos numéricos; por lo que asumiendo en sentido más amplio, es referido de forma aceptado el arte y la ciencia de la reunificación de datos, analizarlos, presentarlos e interpretarlos.

Siendo específicamente en los negocios y en la economía; la información obtenida al reunir datos, analizarlos, presentarlos e interpretarlos pertrecharse ampliamente a directivos, administradores e individuos que deben tomar decisiones a una mejor comprensión del negocio o ambiente económico; lo cual permite que estas sean tomados de forma y manera basándose en mejores informaciones ; su utilización para la decisiones en los negocios y en la economía;

Actualmente en el entorno mundial de los negocios y de la economía, se tiene acceso a enormes cantidades de información estadística; así entonces los directivos y los encargados de la toma de decisiones que acumulan éxitos comprenden la información y dominan usarla eficientemente.

En la extensión del tiempo mayormente, la Estadística ha sido frecuentemente ser odiada por los estudiantes, sin embargo apreciada y reconocida por grandemente profesionistas e investigadores; hoy día es una de las ciencias más útiles e influyentes en la generalidad de los campos del conocimiento, ya que pues ofrece y oferta posibilidades infinitas de desarrollo y aplicación en cualquier entorno.

Con el transcurso del tiempo, cada vez son más recurrentes las diversas aplicaciones de métodos estadísticos en la gestión de proyectos y presupuestos de todo tipo y dimensión; se hace ampliamente explicativo que sanidad y seguridad social son incomprendido sin el empleo de metodologías estadísticas, las cuales son capaces de la recopilación de una proporcionada cantidad de datos de manera continua para su posterior tratamiento mediante potentes paquetes estadísticos informáticos: Actualmente, el análisis estadístico es utilizado para hacer “radiografías” de la situación demográfica y social de un país u otros entornos; además de predicciones de cómo evolucionará (X) población en los próximos 50 o 100 años, igualmente la Estadística es comúnmente empleada por una amplia multitud de profesionales en diversos campos tan como son: La Medicina, La Arquitectura, La Investigación de Mercados, La Meteorología, La Biología, y la Política; se conoce que hoy en día hacer ejercicios indicativos de las preferencias electorales presidenciales en Francia, México o Tejupilco ya generalizándose en otros países del orbe.

1.2. Importancia de la estadística en el campo de la administración

En los momentos actuales en el intervalo de los criterios de la Administración de la Calidad y de la productividad, además de la así aplicación de los diferentes sistemas de gestión de calidad orientados para la toma de decisiones se hace imprescindible la implementación y desarrollo de la Estadística Ya que esta es de esencial importancia en las diferentes empresas, ajustadas desde cualquier área profesional porque ayudan al logro de una adecuada planeación y

control conjugados en los estudios de pronósticos, presupuestos etc. por lo que es concebible que la administración exija la aplicación de la estadística en todas las áreas de los grupo de trabajos administrativos para asegurar la acción de predecir sucesos futuros para así beneficiar a la empresa en la correcta y función de la toma de decisiones, siendo la interrelación acertada entre estas dos grandes ciencias que una complementa a la otra.

La estadística se ha categorizado para los empresarios, como una herramienta que los apoya y ayuda al enfrentamiento y la incertidumbre, teniendo en cuenta el desplazamiento súbito de información disponible, lo cual permite facilitar que los mismos tomen decisiones más inteligentes y directa con mayor rapidez; sirviéndole de apoyo para mantener y asegurar un liderazgo estable ante los trabajadores que dependen de ellos; aunque es de precisión aludir que el desarrollo de las herramientas estadísticas adecuadas en la administración de los negocios está altamente en correlación con la necesidades propias de la gestión; así entonces haremos énfasis en las:

1.3. Aplicaciones de la estadística en la Administración de Negocios

- Estudios de mercado y de opinión.
- Análisis de ratings.
- Procesos de control de calidad.
- Planificación Estratégica.
- Control y evaluación de recursos, etc.

El razonamiento estadístico usa la lógica probabilística en la que, contrariamente a la lógica formal, ninguna proposición puede ser comprobada o rechazada con certeza, sólo con un nivel de significación dado. Mientras la lógica formal es el paradigma metodológico de la matemática y la física clásica, la lógica probabilística lo es de las ciencias factuales, donde la realidad cambia estocásticamente. La incertidumbre respecto al pasado es falta de información, mientras que la incertidumbre respecto al futuro es algo más. La pluralidad de alternativas pasadas puede ser considerada de tipo Laplaciana (todas igualmente posibles) alrededor de la verdadera. Mientras que la pluralidad de posibles futuras alternativas es Darwiniana, lo aleatorio es intrínseco al fenómeno y cambia con

el tiempo. A todo presente le corresponde un gran número de futuros posibles. En realidad, muchos escritores literarios anticiparon con su imaginación creativa lo que los científicos luego formalizaron con nuevos desarrollos teóricos. A este respecto, Jorge Luis Borges escribió: a todo presente se abren infinitos futuros posibles, mucho antes que Illia Prigogine, Premio Nobel de física, desarrollara su famosa teoría de la bifurcación.

En todo paradigma teleológico (finalista) es el futuro que, en cierta medida, determina tanto el presente como el pasado; en el paradigma causal clásico, es el pasado que determina el presente y éste a su vez el futuro; en el paradigma estadístico, el presente es compatible con múltiples futuros y el enlace entre el pasado y el futuro es mediado por el azar. No se pueden prever los acontecimientos, sólo sus probabilidades de presentación.

Resumiendo: En la contemporaneidad es de observación obligada que las estadísticas modernas conjugadas con las herramientas informáticas hacen permisible a los directivos asesores o administradores la manipulación de calidad con la suficiente información para el mejoramiento de los procesos de la empresa; además de encaminar a lograr a las empresas una adecuada y esmerada planeación y control apoyando a los estudios de pronósticos, presupuestos etc. estructura adecuada, revelando la responsabilidad y autoridad de cada una de las partes que integran la organización.

1.4. Clasificación de la estadística

La Estadística Descriptiva y la Estadística Inferencial; son las dos grandes ramas de división de La Estadística La primera aporta las herramientas necesarias que caracterizan a un grupo de elementos o también llamados unidades de análisis sin pretender generalizar los resultados obtenidos a un grupo mayor; objetivo si persigue la estadística inferencial. La Estadística Inferencial es considerada una técnica a través de la cual se avanzan generalizaciones y toma de decisión basadas en la información parcial obtenida en base a técnicas descriptivas; las temáticas más habituales que trata la Estadística Inferencial se encuentran:

- Comparación entre dos o más métodos de trabajo, materiales y produc-

tividad de máquinas, con el fin de elegir cual cumple mejor los requerimientos.

- Comparar instrumentos u equipos de medición para verificar si trabajan con la misma precisión y exactitud.
- Determinar de los factores de un proceso que tienen impacto sobre uno o más características del producto final.
- Encontrar las condiciones de operatividad donde sean reduzcan los defectos o se logre mejor desempeño de proceso.
- Brindar soporte para diseñar productos y procesos robustos a los niveles de ruido que puedan generar variabilidades.

1.5. Objetivo de la estadística como disciplina es brindar soporte en:

- Planificar la búsqueda y obtención de la información.
- Organizar y sistematizar la información para su posterior descripción y análisis.
- Partiendo de la información organizada, efectuar inferencias mediante la estimación y contrastación de la hipótesis.

Es necesidad imperiosa estimar que el modelo de toma de decisión de un tema o problema específico solicita el soporte que contempla la estadística, que imbrica desde la toma de muestra propiamente hasta la confirmación real de la decisión a nivel de confiabilidad definido. Es de suponer que el personal que hace uso frecuente de la estadística debe tener un dominio de la:

1.6. Terminología estadística

- Población: Conjunto de elementos, objetos o unidades de análisis que al menos portan una característica que nos interesa conocer o estudiar.
- Muestra: Una parte cualquiera de la población, un conjunto dado quiera no vacío, o sea, que teóricamente un solo elemento de la población puede ser considerado como una muestra, aunque en la práctica esto ocurra con poca frecuencia.

Los conceptos de muestra y población son conceptos relativos, lo que en un

estudio puede ser una población en otro puede ser una muestra. Los niños de Santa Cruz que asistieron a la escuela en el año 1997 pueden ser parte de una muestra que se haya diseñado para un estudio nacional.

Se considera que una selección de la muestra se puede realizar con la aplicación de técnicas estadísticas obteniendo una muestra probabilística; o simplemente referir la muestra por criterios de expertos. Igualmente, la determinación adecuada de la muestra agrupa el cálculo del tamaño muestral adecuado para el logro de resultados fiables, así como la selección del método de muestreo apropiado.

- **Métodos Muestreo:** El procedimiento para seleccionar la muestra de individuos sobre los que se van recoger los datos debe ser tal que asegure su representatividad. Esto es de primordial importancia para poder llegar a conclusiones que sean generalizable

1.7. Técnicas de muestreo

Las muestras son obtenida son obtenidas de dos tipos: Probabilística y No Probabilística. Las técnicas de muestreo probabilísticas, permiten conocer la probabilidad que cada individuo en estudio tiene de ser incluido en la muestra mediante una selección hecha al azar ; Sin embargo de tipo no probabilísticas, la selección de los sujetos a estudio dependerá de ciertas características, criterios, etc. que él (los) investigador (es) considere (n) en ese momento; por lo que pueden ser poco válidos y confiables o reproducibles; debido a que este tipo de muestras no se ajustan a un fundamento probabilístico, es decir, no dan certeza que cada sujeto a estudio represente a la población blanco. Proponemos entonces varias

1.8. Técnicas de muestreo probabilístico

- **Aleatorio simple:** La misma garantiza que todos los individuos que componen la población blanco tienen la misma oportunidad de ser incluidos en la muestra. Esta significa que la probabilidad de selección de un sujeto a estudio “x” es independiente de la probabilidad que tienen

- el resto de los sujetos que integran forman parte de la población blanco.
- Aleatorio estratificado: Se determina los estratos que conforman la población blanco para seleccionar y extraer de ellos la muestra (se define como estrato a los subgrupos de unidades de análisis que difieren en las características que van a ser analizadas). La base de la estratificación se basa en variable como edad, sexo, nivel socioeconómico, etc. Entonces, se divide la población compuesta por “N” individuos, en “x” subpoblaciones o estratos, con base a variables importantes para la conducción del estudio, y de tamaños respectivos $N_1, N_2, N_3, N_4 \dots, N_k$; y realizando en cada una de estos estratos, muestreos aleatorios simples de tamaño n_i ; para finalmente definir cuantos elementos de la muestra se han de seleccionar de cada uno de los estratos; para lo cual se dispone de las siguientes opciones: asignación proporcional (el tamaño de la muestra de cada estrato es proporcional al tamaño del estrato que le dio origen, respecto a la población total) y asignación óptima (el tamaño de la muestra de cada estrato, son definidos por quien hace el muestreo)
 - Aleatorio sistemático: Cuando el criterio de distribución de los sujetos a estudio en una serie es tal, que los más similares tienden a estar más cercanos. Este tipo de muestreo suele ser más preciso que el aleatorio simple, debido a que recorre la población de forma más uniforme. De este modo, se seleccionará cada hésimo caso.
 - Por conglomerados: Consiste en la elección de forma aleatoria de ciertos barrios o conglomerados ocupados en una región, ciudad, comuna, etc., para luego elegir unidades más pequeñas como cuadras, calles, etc. y finalmente otras más pequeñas, como escuelas, consultorios, hogares (una vez elegido esta unidad, se aplica el instrumento de medición a todos sus integrantes). Si se desea realizar un estudio de prevalencia o una encuesta en habitantes de una localidad, el muestreo aleatorio simple es complejo y de alto costo, ya que estudiar una muestra de tamaño “n”, supone enviar encuestadores a “x” puntos diferentes de la misma; de tal forma que en cada uno de estos puntos, sólo se aplicará una encuesta. Por ello, es que en este tipo de casos se sugiere aplicar muestreo por conglomerados, pues son más económicos y eficientes. En este tipo de muestreo, los sujetos a estudio, se encuentran incluidos en lugares físicos o geográficos (conglomerados); por ende, resulta imprescindible

diferenciar entre sujetos a estudio (quiénes va a ser medidos) y unidad muestral (conglomerado a través del cual se logra acceder a los sujetos a estudio)

- Técnicas de muestreo no probabilístico: Los procedimientos no probabilísticos son aquellos en los que no conocemos la probabilidad de que un elemento de la población pase a formar parte de la muestra ya que la selección de los elementos muestrales dependen en gran medida del criterio o juicio del investigador. La muestra, en este caso, se selecciona mediante procedimientos no aleatorios. Los métodos anteriores (probabilísticos) no son mejores que los no probabilísticos sino que simplemente nos permiten calcular el error muestral que se está cometiendo. Los tipos de muestreo no probabilístico son: muestreo de conveniencia, muestreo discrecional y muestreo por cuotas, de los que pasaremos a explicar brevemente:
- Muestreo de conveniencia: El investigador decide qué individuos de la población pasan a formar parte de la muestra en función de la disponibilidad de los mismos (proximidad con el investigador, amistad, etc.).
- Muestreo discrecional.: La selección de los individuos de la muestra es realizada por un experto que indica al investigador qué individuos de la población son los que más pueden contribuir al estudio. Este muestreo es adecuado si dentro de la población que queremos estudiar, existen individuos que no queremos que se nos escapen por utilizar un método totalmente aleatorio o de conveniencia
- Muestreo por cuotas: Si se conocen las características de la población a estudiar, se elegirán los individuos respetando siempre ciertas cuotas por edad, género, zona de residencia, entre otras que habrán sido prefijadas.
- Intencional: Permite seleccionar casos característicos de una población limitando la muestra sólo a estos casos. Se utiliza en escenarios en las que la población es muy variable y consiguientemente la muestra es muy pequeña. Por ejemplo, entre todos los sujetos con CA, seleccionar a aquellos que más convengan al equipo investigador, para conducir la investigación.
- Accidental o consecutivo: Se fundamenta en reclutar casos hasta que se completa el número de sujetos necesario para completar el tamaño

1.9. Fortalezas y Debilidades de la Aplicación de Técnicas de Muestreo

Se puede comentar que cada técnica de muestreo tiene sus ventajas y desventajas; sin embargo, en términos generales existen ventajas y desventajas comunes para todas ellas:

1.10. Fortalezas

Reducción de costos: Los costes de un estudio serán menores si los datos de interés se pueden obtener a partir de una muestra de la población blanco. Por ejemplo, cuando se realizan estudios de prevalencia de un evento de interés, es más económico medir una muestra representativa de 1500 sujetos de la población blanco, que a los 250.000 individuos que la componen.

Eficiencia: Al trabajar con un número reducido de sujetos a estudio, representativos de la población blanco; el tiempo necesario para conducir el estudio y obtener resultados y conclusiones será ostensiblemente menor.

1.11. Debilidades

Inadecuada representación de la población blanco: Esto puede ocurrir si se decide trabajar con muestras pequeñas.

Unidad de análisis

Es relativamente ambiguo el concepto de U.A., ya que las definiciones abrigadas por distintos metodólogos no suelen ser consistentes entre sí.

- a). Si hablamos de unidad es porque nos referimos a un dominio circunscripto y diferenciable con propiedades inherentes. Dominio también delimitado, en tanto podemos trazar una especie de frontera que individualice una totalidad y la distinga de otras entidades.

El conjunto de entidades y relaciones que hemos circunscripto adquiere así el estatuto de unidad u organización diferenciada.

- b). Si hablamos de análisis es porque suponemos que la unidad definida es pasible de conocerse siguiendo algún tipo de procedimiento de indagación.

Es decir que, al pretender analizar una unidad, estamos suponiendo que ésta es inteligible y que para lograr conocer algo de ella debemos aplicar determinados procedimientos. Las U.A. pueden constituirse y/o reformularse en distintos momentos del proceso que se va abordar. Lo que sí pretendemos decir es que una U.A. sólo puede formularse en un proceso de investigación cuando ya han sido formulados el tema, los problemas e interrogantes centrales. Esto no significa que, una vez establecida una U.A., no pudieran aparecer otros problemas o interrogantes (situación bastante frecuente), sino que no es posible hablar de U.A. sin tema-problema previo.

En el orden de las ideas anteriores las unidades de análisis: se conceptualizan como los elementos de la población objeto de estudio. Pueden ser de variada naturaleza en dependencia de los objetivos y del fenómeno estudiado.

Después de lo anteriormente expuesto la unidad de análisis son los elementos en los que recae la obtención de información y que deben de ser definidos con propiedad, es decir precisar, a quien o a quienes se va a aplicar la muestra para efectos de obtener la información.

Variables

Concepto de trascendencia para la estadística, pues constituye el punto de partida para definir métodos de representación a utilizar, medidas de resumen a calcular y que técnicas de análisis se pueden aplicar.

Es referido como variable alguna característica de la población que puede asumir diferentes comportamientos, valores, o grados de intensidad entre los diferentes elementos individuos o unidades de análisis que la conforman. Por ejemplo la edad es una característica que asume diferentes valores de un individuo a otro, el sexo, el estado civil, los servicios de un hospital, estas son consideradas variables son todas variables.

Resumiendo toda característica asociada a los objetos y fenómenos que nos rodean constituyen variables: los colores de las cosas, la estatura de las personas, la altura de las edificaciones, el volumen de los recipientes, la cantidad de autos que pasan la noche en un parqueo, o los países ganadores en las Copas Mundiales de Boxeo, entre otras. Pero, observese que no son variables por el simple hecho de ser características, sino porque pueden asumir valores diferentes en dependencia de que objeto o fenómeno midas, además, Estas no se miden todas de la misma manera, unas se miden numéricamente y otras no, de aquí que las variables pueden clasificarse en dos grandes grupos: cuantitativas y cualitativas, respectivamente

- Las variables cualitativas: son las que las diferencias entre un elemento y otro son atributos, cualidades no medibles en términos numéricos; clasificándose en nominales u ordinales. Nominales: No existen diferencias de magnitud o intensidad entre sus categorías Ordinales: Tienen implícito diferencias de magnitud o intensidad entre sus categorías que les confiere cierto orden
- Las variables cuantitativas son: las que las diferencias que existen existentes entre los diferentes elementos de la población dada son medibles numéricamente; clasificándose a su vez en discreta o discontinua y continuas
- Discreta o discontinua : Toman valores siempre enteros de manera que entre dos valores enteros consecutivos no existen posibilidades practicas ni teóricas de que hallan valores intermedios
- Continuas: Entre dos valores enteros consecutivos existen infinitos valores intermedios, aunque frecuentemente se empleen valores enteros.

Resulta muy importante la identificación del tipo de variables que estudiamos en las unidades de análisis de una investigación, ya que todo el procesamiento estadístico de los datos depende del tipo de variable.

1.12. Presentación de datos estadísticos

La presentación de resultados se inicia a partir de lo establecido en el diseño conceptual y es hasta que se obtiene la información en la etapa de procesa-

miento que se determinan los medios y las formas para presentarla o publicarla, particularmente, la elaboración de cuadros y gráficas es una herramienta para la presentación de la información.

En los análisis estadísticos, es frecuente utilizar representaciones visuales complementarias de las tablas que resumen los datos de estudio. Con estas representaciones, adaptadas en cada caso a la finalidad informativa que se persigue, se transmiten los resultados de los análisis de forma rápida, directa y comprensible para un conjunto amplio de personas.

Componentes de un cuadro:

Un cuadro estadístico, también llamado tabulado o tabulación, consiste en una presentación ordenada de una agrupación de datos cuantitativos, sea en una sola columna o en un solo renglón como también, en columnas o renglones cruzados. Este incluye los valores numéricos, así como las descripciones conceptuales las que hacen referencia los mismos; además, incorpora indicaciones muy particulares acerca del origen de los datos, además aclaraciones específicas que son útiles para el usuario que esté interesado en el conocimiento de aspectos de naturaleza conceptual, técnica o metodológica. Estos cuadros requieren indicaciones sobre su modulación o presentación en distintas páginas, como también sobre su identificación numérica cuando se trata de un conjunto de cuadros.

Un cuadro incluye seis componentes:

1.13. Título

Descripción del contenido del cuadro, basado en la indicación de cuatro aspectos:

- **Enunciado:** Expresa el nombre de las categorías, indicadores o variables.
- **Cobertura geográfica:** Indica el espacio territorial al que se refiere el conjunto de datos estadísticos.

- Referencia temporal: Indica la fecha o periodo al que se refieren los datos estadísticos.
- Unidad de medida: Indica la referencia de una magnitud constante adoptada.

Columna matriz

- Enuncia las categorías, variables, clasificaciones o indicadores referidos a los datos, según su agrupamiento en renglones
- La columna matriz presenta dos apartados:
- B1 Encabezado de la columna matriz; Enuncia las categorías, variables o indicadores que están descriptores de renglón
- B2 Descriptores de renglón. Enuncian los conceptos a los que se refieren los datos de cada renglón.

Encabezado en columnas de la matriz de cifras:

- En esta son enunciados las categorías, variables, clasificaciones o indicadores de los descriptores de columna; los mismos son los conceptos a que se refieren los datos de cada columna.

Matriz de cifras:

- Espacio destinado a la incorporación de los valores cuantitativos de los datos, o los símbolos que explican la ausencia de éstos.

Pies de cuadro

- Precisiones conceptuales, técnicas o metodológicas que facilitan la interpretación y comprensión de los datos presentados:
- E1 Nota; Información general sobre definiciones o referencias técnicas o metodológicas de los datos estadísticos presentados en cada cuadro.
- E2 Llamadas; Información específica aplicable a determinada parte o elemento del cuadro.
- E3 Símbolos aclaratorios. Todos los signos convencionales que se indi-

can mediante una letra o una abreviatura en el contexto del cuadro estadístico. E4 Fuentes. Indica la procedencia de la información contenida en el cuadro estadístico Indicaciones de orden:

- Expresan el número de cuadro y la secuencia de las partes cuando éste es fraccionado, además permite ver la secuencia de la información presentada y la interrelación que se pudiera establecer entre ellas; En los casos de preparación y presentación de cuadros para internet o en disco compacto, pudiera no necesitarse el elemento de secuencia, debido a que existe la posibilidad de integrar toda la información completa ejemplo; en una página de Excel; la sábana íntegra con toda la información de un cuadro.
- F1 Número de cuadro. Expresa la ubicación ordinal correspondiente al cuadro en el documento; el número ha de estar antecedido por la palabra “Cuadro”.
- F2 Indicaciones de continuidad. Especifica la secuencia de las partes en las cuales se fracciona el cuadro, cuando no es suficiente el espacio para presentarlo en una sola página, como es el caso de las publicaciones impresas.

Consideraciones:

De carácter general: Es necesario cuidar la claridad de los enunciados que describen el contenido del cuadro. En ese sentido, las recomendaciones generales son:

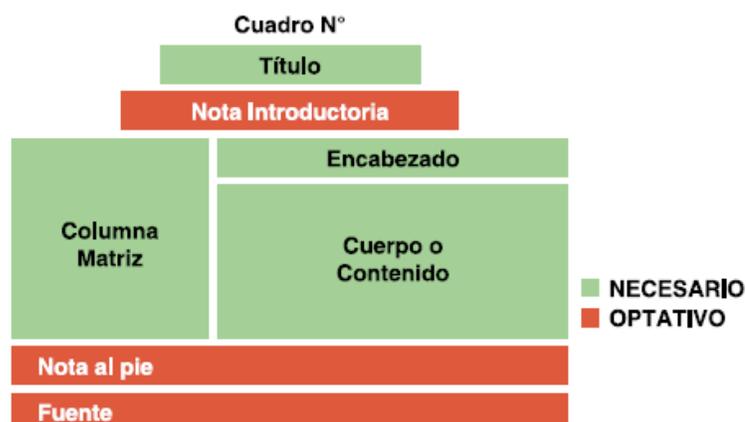
- Utilizar un lenguaje preciso y claro en todos los enunciados que describen los títulos, descriptores de columna y renglón, notas, llamadas y fuentes.
- Escribir los textos con mayúsculas y minúsculas en los diversos componentes del cuadro (título, encabezado de columna matriz, descriptores de renglón y encabezado de matriz de cifras). El uso de mayúsculas en estos elementos estará reservada a la inicial del enunciado y a los nombres propios. En los casos de las palabras Nota, Fuente, Cuadro y Continúa, solamente la letra inicial irá con mayúscula.
- Vigilar el número y diversidad de categorías, indicadores y variables por

incorporar al diseñar un cuadro, evitando en lo posible una excesiva heterogeneidad en la combinación de conceptos. Se recomienda, además, que en un cuadro no se cruce la información de más de tres variables. Debe considerarse, también, la extensión de las clasificaciones que contenga cada una de las variables.

- En cuadros impresos, ordenar y distribuir los conceptos en las columnas y renglones de tal manera que se facilite la identificación de conceptos específicos y la localización de los datos correspondientes, destinando los renglones para los desgloses más amplios y las descripciones más largas, y las columnas para desgloses y enunciados más cortos, evitando en lo posible grandes crecimientos en sentido vertical y horizontal. En el caso de cuadros en medios informáticos (como las hojas de Excel) se tiene la libertad de extenderse e integrar el cuadro completo.

Incluir, como mínimo, los siguientes cinco componentes del cuadro estadístico, descritos en los componentes de un cuadro:

- **Título:** Es importante que el título del cuadro permita al usuario conocer lo más detalladamente posible el contenido del mismo, de tal forma que el índice o relación de cuadros sea una herramienta que facilite al usuario la localización de la información que requiere.
- **Columna matriz:** La columna matriz está formada por dos partes: el encabezado, en el que se enuncian las categorías, variables o indicadores que aparecen en los descriptores de renglón; y los descriptores de renglón, en los que se enuncian los conceptos a los que se refieren los datos de cada renglón.
- **Encabezado de la matriz de cifras:** El enunciado del encabezado de la columna matriz debe ir en **BOLD** y estar alineado a la izquierda (respecto al espacio asignado a ese encabezado) y centrado respecto a las placas superior e inferior.
- **Matriz de cifras:** Es la parte del cuadro estadístico donde se presentan los elementos numéricos, en correspondencia con los descriptores del renglón y la columna.
- **Fuente:** El enunciado de fuentes permite a los usuarios del cuadro conocer la procedencia de la información



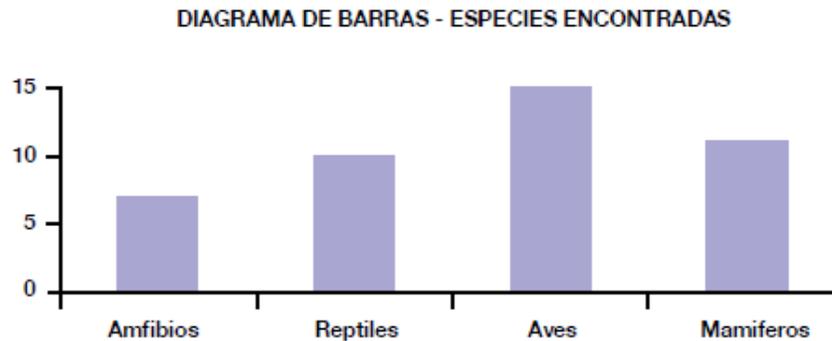
Gráficos

También llamados “figuras”. Los gráficos estadísticos son herramientas que ayudan a representar visualmente los datos estadísticos. Existen distintos tipos de gráficos, los cuales se elaboran de acuerdo con el tipo de información que se desee mostrar. Existen diversos tipos de gráficos que se utilizan de acuerdo con el tipo de información que se quiere presentar.

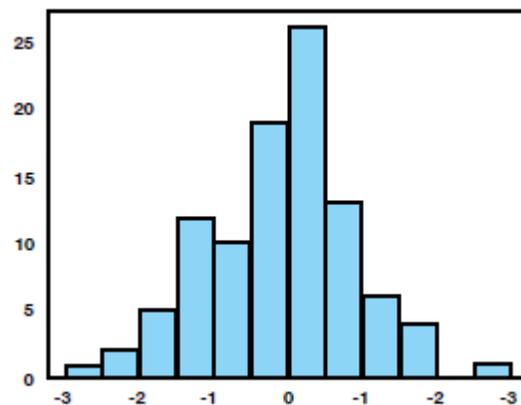
Tipos de representaciones graficas:

Cuando se muestran los datos estadísticos a través de representaciones gráficas, se ha de adaptar el contenido a la información visual que se pretende transmitir. Para ello, se barajan múltiples formas de representación:

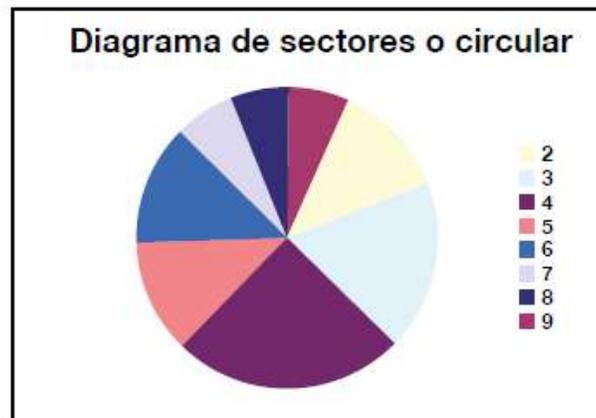
- Diagramas de barras: muestran los valores de las frecuencias absolutas sobre un sistema de ejes cartesianos, cuando la variable es discreta o cualitativa.



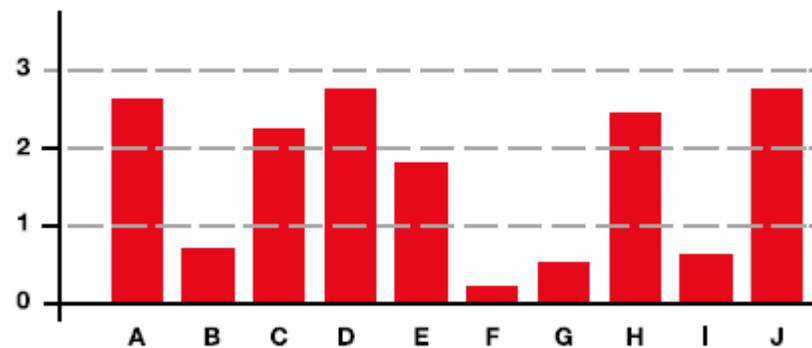
- Histogramas: formas especiales de diagramas de barras para distribuciones cuantitativas continuas.



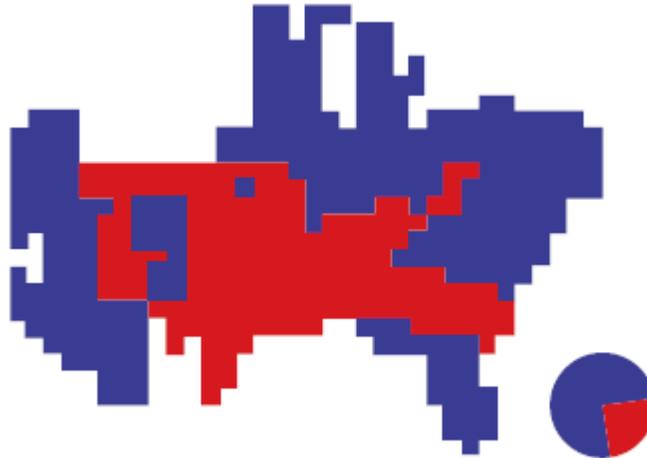
- Polígonos de frecuencias: formados por líneas poligonales abiertas sobre un sistema de ejes cartesianos.
- Gráficos de sectores: circulares o de tarta, dividen un círculo en porciones proporcionales según el valor de las frecuencias relativas.



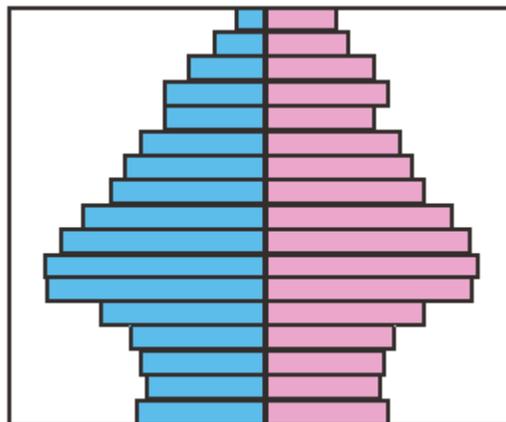
- Pictogramas: o representaciones visuales figurativas. En realidad son diagramas de barras en los que las barras se sustituyen con dibujos alusivos a la variable.



- Cartogramas: expresiones gráficas a modo de mapa.



- Pirámides de población: para clasificaciones de grupos de población por sexo y edad.



Los diagramas de barras se usan para representar gráficamente series estadísticas de valores en un sistema de ejes cartesianos, de manera que en las abscisas se indica el valor de la variable estadística y en las ordenadas se señala su frecuencia absoluta.

Estos gráficos se usan en representación de caracteres cualitativos y cuantitativos discretos. En variables cuantitativas continuas, se emplea una variante de

los mismos llamada histograma.

Polígonos de frecuencias: Para construir polígonos de frecuencias, se trazan las frecuencias absolutas o relativas de los valores de la variable en un sistema de ejes cartesianos y se unen los puntos resultantes mediante trazos rectos. Con ello se obtiene una forma de línea poligonal abierta.

Los polígonos de frecuencias se utilizan preferentemente en la presentación de caracteres cuantitativos, y tienen especial interés cuando se indican frecuencias acumulativas. Se usan en la expresión de fenómenos que varían con el tiempo, como la densidad de población, el precio o la temperatura.

Gráficos de sectores: En los diagramas de sectores, también llamados circulares o de tarta, se muestra el valor de la frecuencia de la variable señalada como un sector circular dentro de un círculo completo. Por ello, resultan útiles particularmente para mostrar comparaciones entre datos, sobre todo en forma de frecuencias relativas de las variables expresadas en forma de porcentaje.

Pictogramas y cartogramas: Para aligerar la presentación de datos estadísticos, con frecuencia se recurre a imágenes pictóricas representativas del valor de las variables. Dos formas comunes de expresión gráfica de los datos son:

- Los pictogramas, que muestran diagramas figurativos con figuras o motivos que aluden a la distribución estadística analizada (por ejemplo, una imagen antropomórfica para indicar tamaños, alturas u otros).
- Los cartogramas, basados en mapas geográficos que utilizan distintas tramas, colores o intensidades para remarcar las diferencias entre los datos.
- Mapa estadístico, los mapas, por definición muestran la localización de los datos. Si su único objetivo es mostrar dicha localización sin adjudicar un valor relativo o absoluto a los datos, es decir, el nivel de medición es nominativo, se denominan mapas cualitativos, no obstante la mayoría de los mapas estadísticos además cuantifican el valor de los datos, y estos se denominan mapas cuantitativos

Pirámide de población: Otra forma corriente de presentación visual de datos

estadísticos es la llamada pirámide de población.

Las pirámides de población se utilizan en la expresión de informaciones demográficas, económicas o sociales, y en ellas se clasifican comúnmente los datos de la población del grupo de muestra considerado en diferentes escalas de edad y diferenciada por sexo.

El cuerpo del gráfico varía dependiendo del tipo de gráfico que se utiliza, sin embargo, la mayoría tiene características en común, tales como:

- Eje de valores y título del eje (excepto gráficos circulares y radiales). En el caso de los gráficos de dispersión, los dos ejes son ejes de valores.

Cuando se trata de una variable continua de razón, en el cual el valor mínimo posible es cero, por ejemplo: porcentajes e ingresos en miles de colones; el eje siempre debe iniciar en cero. En caso de que esto no sea posible, debe insertarse una marca de “ruptura” o “corte” al inicio del eje.

- Eje de categorías y título del eje (excepto gráficos circulares, radiales y gráficos de dispersión).
- Leyenda: Indica la correspondencia del color o trama utilizados en el gráfico con la categoría o variable respectiva.
- Líneas de división Son importantes para guiar y ubicar al lector de una mejor manera en los valores respectivos. Las líneas de división verticales se utilizan exclusivamente para gráficos de barras horizontales.

En el caso de los gráficos de barras verticales, lineales o de dispersión, únicamente se deben utilizar líneas de división horizontales. Además, se recomienda que el color de la línea sea gris claro para evitar que el gráfico luzca saturado.

- Etiquetas Las etiquetas son los datos con los que se construye el gráfico. Es posible mostrarlas u ocultarlas y esta decisión depende de la cantidad de valores que contenga el gráfico. En ocasiones no son necesarias porque con solo las líneas de división, el lector logra ubicar fácilmente el valor correspondiente en el eje.

Por el contrario, cuando se tienen muchos valores (etiquetas) y es indispensable mostrarlos, se recomienda la implementación de un cuadro estadístico, en lugar de un gráfico.

Consideraciones importantes:

- Cuando los gráficos están saturados, es decir, cuando se introducen muchas variables o categorías, pierden su objetivo, el cual es representar información de manera resumida y fácil de comprender visualmente.
- Deben ser claros, ya que se espera que se expliquen por ellos mismos, sin necesidad de recurrir a textos adicionales para su comprensión.
- Deben ser visualmente proporcionales debido a que, mediante las dimensiones de las barras o de las áreas del gráfico, se representan las magnitudes reales.
- Cuando la escala del gráfico no inicia en cero debe hacerse una aclaración en el texto y señalarlo en el gráfico.
- Se debe agregar los valores o las cifras en el gráfico, esto con el objetivo de que el usuario tenga la posibilidad de conocer el valor exacto que se está representando; sin embargo, cuando el gráfico tiene muchos valores su lectura puede complicarse. Al tomar esto en consideración, es posible agregar el valor o la cifra exacta, únicamente en los puntos más relevantes del gráfico.
- El rango de las escalas no se debe manipular para sobre dimensionar o minimizar el efecto de los datos representados, ello distorsiona la percepción y por ende su interpretación.
- Una alternativa para el uso de colores en los gráficos es utilizar tramas. Ya sea que se utilicen colores o tramas, se recomienda que cada categoría o variable mostrada tenga un color o trama distinto a las demás, con el fin de evitar confusiones al lector.

Método de recolección de datos

Toda medición o instrumento de recolección de los datos debe reunir dos elementos esenciales: confiabilidad y validez.

Confiabilidad: la confiabilidad de un instrumento de medición se refiere al grado en que su aplicación repetida al mismo sujeto u objeto produce iguales resultados **Validez:** en términos generales se refiere al grado en que un instrumento realmente mide la variable que pretende medir.

La validez es un concepto de cual puede tenerse diferente tipo de evidencia: evidencia relacionada con el contenido, evidencia relacionada con el criterio, evidencia relacionada con el constructo.

Factores que pueden afectar la confiabilidad y la validez.

- La improvisación
- Instrumentos no validados
- Instrumentos que resultan inadecuados a las personas a las que se les aplica
- Condiciones en que se aplica el instrumento.

Existen diversas técnicas que posibilitan la recolección de información como:

La encuesta

El método de encuesta se basa en un conjunto de preguntas que se le formulan al participante, cuya información constituye la información primaria necesaria para el investigador acorde con los objetivos de su investigación. Esta forma de coleccionar la información primaria puede ser por preguntas orales o escritas y puede tener diversas formas.

La entrevista

Es el medio que permite la obtención de información de fuente primaria, amplia y abierta, en dependencia de la relación entrevistador entrevistado. Para ello es necesario que el entrevistador tenga definido claramente los objetivos de la entrevista y cuáles son los aspectos relevantes sobre los que se pretende obtener información. La conducción de la entrevista debe ser precisa y dinámica, en

un clima de sinceridad, cordialidad y confianza. Para ello se requiere que el entrevistador tenga un buen adiestramiento. Existen distintos tipos de entrevista:

Según la relación que se establece con el investigador Cara a cara: cuando se realiza frente al encuestado, tiene la ventaja que permite observar las reacciones, gestualidad y formas del mismo. Esto enriquece la información.

En profundidad: las preguntas están dirigidas a profundizar en aspectos previamente identificados en el proceso de la investigación, requiere de un vínculo muy estrecho entre el sujeto investigador y el sujeto participante.

Según la forma que adopta la entrevista: Estandarizada: las preguntas y su secuencia se sigue en un mismo orden para todos los entrevistados. No estandarizada: a diferencia de la anterior se le entrega al entrevistado un tema o conjunto de temas para que lo desarrolle.

Semi-estandarizada: se combinan las preguntas previamente elaboradas y estandarizadas, con la forma no estandarizada, con la suficiente flexibilidad para que el entrevistado lo desarrolle.

El cuestionario

El cuestionario es otra forma de obtener información primaria, por sus características es aplicable a escala masiva. Está destinado a obtener respuestas a preguntas previamente elaboradas que responden al problema de investigación. Se utiliza para ello un formulario impreso. Existen requisitos básicos para su confección como son: llamado a la cooperación, elaboración de preguntas claras, precisas, elaborar las preguntas sin que induzcan las repuestas, preguntas que permitan la validación de las respuestas correspondientes a las preguntas principales (preguntas muelle o colchón).

Tipos de cuestionario: Por correo: formulario impreso, que se envía por correo a la población objeto de estudio. Grupal: es aplicado a un grupo, previo acuerdo del lugar y hora, en presencia del investigador, que realiza la demanda de cooperación, formula las instrucciones y aclaraciones pertinentes.

Las encuestas y los cuestionarios son con frecuencia utilizados para recoger información que se procesa por técnicas cuantitativas.

Entre las técnicas más utilizadas para recoger información de tipo cualitativa, dirigida a obtener ricas descripciones y explicaciones sobre los procesos en contextos locales, narradas en las propias palabras de los actores sociales se encuentran:

Informantes clave

Los informantes clave son aquellos individuos que poseen conocimientos, posición dentro del grupo, o destrezas comunicativas especiales y que están dispuestos a cooperar con el investigador.

La selección de los informantes clave dependerá de la información que se pretenda obtener según los objetivos del estudio. En el caso de la investigación de la salud pueden constituir informantes claves los protagonistas de la acción de salud: médicos/as, enfermeras/os, promotoras/es, etc., otros actores sociales que intervienen directa o indirectamente en las acciones de salud como, profesores y alumnos de escuelas, comerciantes, vecinos que hayan permanecido por largo tiempo en el área objeto de estudio; miembros de instituciones comunitarias y afines, etc. y aquellos que sean conocedores de los ideales del grupo objeto de estudio. En estos casos pueden ser verdaderos líderes (formales, cuando son representativos de espacios institucionales o políticos de poder asignados dentro del grupo; informales, cuando los espacios de poder son asignados por el propio grupo).

Por lo general, los informantes claves, son individuos atípicos, por lo que deben ser seleccionados con cuidado, si se pretende que tengan un nivel adecuado de representatividad respecto al grupo.

La selección de los informantes clave puede hacerse:

Inducida: definida por la intención de la información a obtener. Por ejemplo, personas que tengan una conducta no saludable, trabajadores indisciplinados

que participan en determinados disturbios y problemas disciplinarios en el área de trabajo (para identificar causas de antagonismos en los grupos, exclusiones y marginaciones); funcionarios involucrados en la organización de historias clínicas y otros documentos de control (para opinar sobre discriminación de trabajadores relevantes o proponer posibles causas de las indisciplinas).

Al azar: de un grupo de informantes claves identificados, se utiliza un medio de selección aleatoria para su incorporación al estudio.

Bola de nieve: Se selecciona un informante clave, y ese señala dos y estos dos señalan dos más, así se continúa hasta alcanzar los definidos como necesarios para el estudio. Este tipo de selección se utiliza mucho cuando se requiere información muy privada, tal es el caso de los estudios de violencia intrafamiliar.

Historias de vida La historia de vida es un registro pormenorizado de un sujeto seleccionado que permite el análisis e interpretación de una o varias categorías u objetos de análisis que constituyen el objeto de la investigación. Para la historia de vida se seleccionan personas que tengan una historia que permita reconstruir procesos.

Otras formas de recolección de información:

- **Vía Internet:** se realiza por este medio de comunicación
- **Telefónica:** se realiza por este medio de comunicación
- **Observación directa:** Es uno de los primeros métodos científicos utilizados en la investigación y se utiliza para la obtención de información primaria acerca de los objetos investigados o para la comprobación empírica de las hipótesis. La observación científica es sistemática, consciente y objetiva. Su valor radica en que permite obtener la información del comportamiento del objeto de investigación tal y como este se da en la realidad, es decir, información directa, inmediata sobre el fenómeno u objeto investigado. Para lograr una buena calidad en la observación resulta determinante establecer una adecuada definición de las situacio-

nes, fenómenos o estados objeto de la observación.

- **Técnicas grupales** Las técnicas grupales son aquellas que se realizan con grupos de personas que pueden ofrecer información sobre el asunto o fenómeno objeto de estudio. Entre las mismas se pueden mencionar: las asambleas comunitarias, los grupos de discusión, también llamados focales, nominales.

Diversas son las técnicas que se pueden utilizar para el trabajo en grupo, en todos los casos deben corresponderse a los objetivos del estudio, por citar algunas: brainstorming (lluvia de ideas), brainswriting (lluvia de escritura), técnicas de expertos por validación de consenso, sociodrama, pictograma. etc.

Entre las técnicas de carácter grupal más mencionadas se señalan: el grupo focal y el grupo nominal, a continuación ofrecemos una breve caracterización de las mismas.

El grupo focal: se forma con el objetivo de obtener información sobre percepciones, creencias, lenguaje. Su estructura se compone por 8 o 10 personas seleccionadas acorde con la estrategia trazada para cumplir los objetivos del estudio. Lo dirige un/a facilitador/a que guía la orientación de la guía de preguntas, debe también designarse un observador/a que registre todo el proceso. Lo ideal para este método es filmar y grabar en video; siempre se debe registrar cuidadosamente la información. Para conducir al grupo en cada incorporación se realiza una de las interrogantes y después se incorpora una de cada vez, profundizando para obtener información más útil.

A partir de la información que se obtiene se pueden formular hipótesis y se puede delimitar problemas para estudios de mayor alcance. Utilizando esta técnica se pueden conocer ideas, percepciones, actitudes, en otro orden puede facilitar la interpretación de investigaciones cuantitativas.

La técnica del grupo nominal

El objetivo del uso de esta técnica está dirigida a evaluar la percepción comunitaria de los problemas, esta herramienta permite superar la desigualdad en

la representación de opiniones. En este tipo de grupo el número de personas no debe exceder a 7. En este caso el/la moderador/a dirige el trabajo del grupo, el moderador hace la pregunta al grupo; posteriormente cada participante escribe en una hoja la respuesta (el tiempo para este ejercicio debe ser aproximadamente de 15 minutos).

En un segundo momento cada participante lee su respuesta, en la medida que cada uno la lee, el líder del grupo la escribe en la pizarra o en un rotafolio y le asigna un número a cada respuesta, se procede de esta manera hasta que se tengan todas las contribuciones. Esta forma garantiza que todos los integrantes del grupo puedan participar plenamente.

Posteriormente, el moderador preguntará si cada pregunta se entiende, aclarará el significado de cada respuesta. En este momento sólo se realizan aclaraciones no se argumenta.

Se realiza una votación preliminar, a partir de los elementos propuestos que consideren más importantes, esta operación permite ordenar la secuencia de las respuestas según su importancia.

El último paso es ordenar jerárquicamente y realizar una votación final que permite ponderar un orden a las respuestas.

Registro continuo: Consiste en anotar la/s pauta/s de comportamiento indicando el momento en el que aparecen y su duración, midiendo frecuencias de aparición, duración total de la pauta o latencia.

Mediante esta técnica obtenemos un registro exacto y fiel del comportamiento y es útil para medir secuencias de comportamiento. Sin embargo, su utilización puede estar limitada por razones prácticas ya que supone un mayor esfuerzo para el observador que con el registro temporal y como consecuencia se pueden recoger menor número de pautas de forma simultánea. En resumen utilizar una técnica u otra, cuantitativa o cualitativa, dependerá de los objetivos del estudio y de su alcance. Las técnicas de recolección de datos utilizadas en la investigación cualitativa permiten recoger las experiencias de los actores en

sus propias palabras, sus propios ritmos, y que mejor recojan sus visiones del mundo. En general estas técnicas permiten una mejor aprehensión de las dinámicas particulares de la experiencia humana, de los procesos vividos y de la construcción de las representaciones sociales.

En general el objetivo central del análisis cualitativo no es la generalización estadística sino la aprehensión de las particularidades y significados de los casos dentro de los contextos de la selección, la selección de muestras por medio de técnicas aleatorias no es apropiada y muchas veces ni siquiera posible por no contarse con el marco muestral adecuado.

Etapas de la investigación estadística:

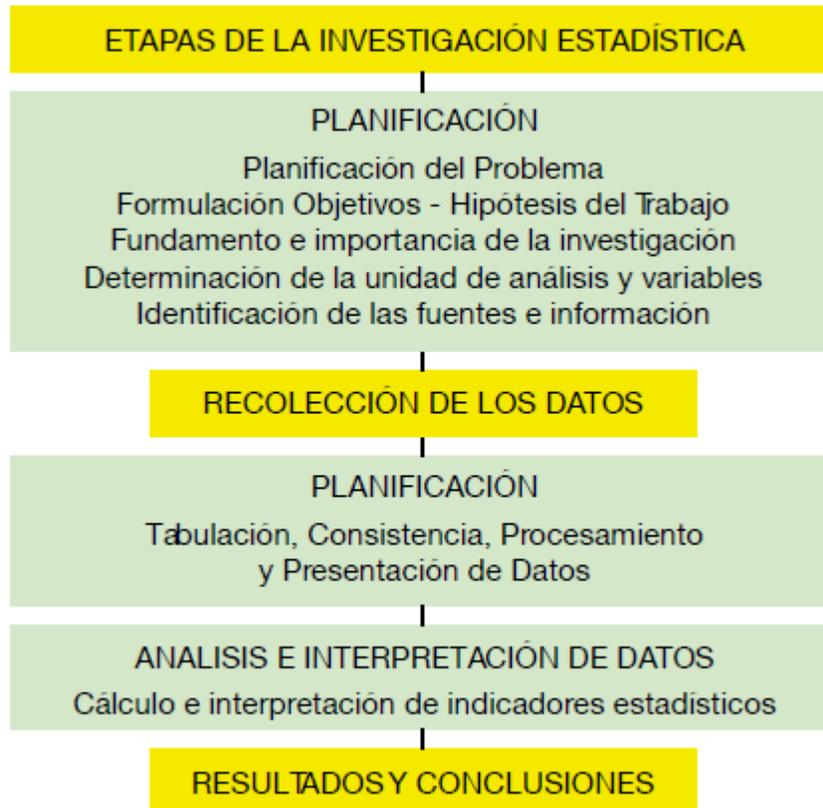
Las actividades que involucra una investigación estadística pueden clasificarse con arreglo a diversos criterios. Uno de ellos se relaciona con el orden cronológico en que deben ser realizadas. En este sentido, puede hablarse de cinco grandes etapas o fases:

Planificación

La planificación no se realizará adecuadamente si antes no se ha definido claramente la naturaleza y objetivos de la investigación así como la evaluación de los conocimientos que se tienen sobre el problema y de las hipótesis que se han formulado para explicarlo.

Recolección de datos

En segundo lugar, se considera la ejecución de la investigación, la misma que implica la recolección, clasificación y análisis de la información recogida según lo planificado.



Los principales puntos que deben considerarse al recoger la información son:

- Los errores que pueden cometerse en la recolección de los datos.
- Las ventajas y limitaciones de los diversos métodos empleados en la recolección de la información.
- Las condiciones que deben reunir los individuos que se estudian y los procedimientos más pertinentes para su elección.

Etapa que implica la revisión cuidadosa de la información recogida para resumirla y presentarla convenientemente.

Se consideran los siguientes aspectos:

- Revisión y corrección de la información recogida etapa labor es llamada: Consistencia.

- Presentación de la información mediante cuadros, tablas y gráficos.
- Análisis e Interpretación de Resultados.

Se interpreta y compara los resultados de los indicadores estadísticos o estadígrafos. Si el estudio fue realizado conforme a lo que se había planificado y con los resultados a la vista se concluirá si las hipótesis ha sido verificadas o no, proponiéndose las recomendaciones pertinentes.

Resultados y Conclusiones

Finalmente exponer los principales resultados de acuerdo a los objetivos. Indicar lo más importante, si se acepta o se rechaza las hipótesis.

Concepto de Censo

Es la principal fuente de datos demográficos por la gran cantidad de información que maneja: da una fotografía de la población, describe estadísticamente las poblaciones humanas consideradas desde un punto de vista cuantitativo. Es la fuente primaria de las estadísticas básicas de población que son necesarias para fines gubernamentales y aspectos de planificación económica y social. Se usa:

1. Como base para el análisis y la evaluación demográfica.
2. Para proyectar, establecer y desarrollar políticas de gobierno.
3. Para hacer estimaciones de las distintas variables captadas en el Censo.
4. Como “marco muestral” para encuestas.
5. Como referencia para las estadísticas continuas (vitalales).
6. Para determinar los sistemas electorales de un país.

Fortalezas de Censo

- La información obtenida puede presentarse por unidades administrativas u otras unidades de estratificación cualquiera sea su tamaño, pudiendo obtener datos para áreas pequeñas.
- Punto de referencia para las estadísticas continuas.
- Único procedimiento utilizable para saber sobre fenómenos con poca

frecuencia.

Debilidades del censo

- Alto coste (humano y material) dado que exige el empleo de una gran cantidad de recursos de personal, financieros y materiales.
- Es necesaria una vasta organización que abarque todo el universo a investigar, procurando evitar omisiones y duplicaciones.
- Demora en la obtención de resultados.
- En algunos casos, la información que se obtiene puede ser de inferior calidad (mayores errores) a la que se obtendría si la investigación se realizara por muestreo.

Tabla de distribución de frecuencia

A partir de las tablas de distribución de frecuencias, se proporcionará un conjunto de herramientas o técnicas estadísticas para el tratamiento de la información, tanto cualitativa como cuantitativa, que facilitan el manejo y análisis de los datos.

La distribución de frecuencias es la representación estructurada, en forma de tabla, de toda la información que se ha recogido sobre la variable que se estudia.

Toda variable o atributo tiene asociada una distribución de frecuencias, que implica una correspondencia biunívoca entre dos conjuntos. Por un lado, tenemos el conjunto formado por los valores que toma la variable o el atributo y por el otro el de las frecuencias relacionadas con ellos. La asociación existe en ambos sentidos, pues a cada valor de la variable o atributo, denominado clase, le corresponde una frecuencia y a su vez cada frecuencia se identifica con una clase de la variable.

Frecuencia absoluta y relativa

Frecuencia absoluta: Es la cantidad de datos que integran cada una de las clases, o sea que son las repeticiones que encontramos dentro de una misma cla-

se. La correspondencia entre los valores de la variable y su frecuencia absoluta determina lo que denominamos “distribución de frecuencias absolutas”.

Frecuencia relativa: Es la cantidad de repeticiones obtenidas para cada clase, en relación al total de las observaciones. Resulta de dividir la cantidad de elementos de cada clase (frecuencia absoluta) por el tamaño de la muestra. Las frecuencias relativas son un porcentaje, ya que relacionan una parte del conjunto con el total. La correspondencia entre los valores de la variable y su frecuencia relativa determina lo que denominamos “distribución de frecuencias relativas”.

Ejemplificamos

Edad años	Frecuencia absoluta	Frecuencia relativa	%	Frecuencia acumuladas	
				Absoluta	Relativa
20-24	20	0.29	29	20	29
25-29	18	0.27	27	38	56
30-34	15	0.22	22	53	78
35-39	10	0.15	15	63	93
40-44	5	0.07	7	68	100
Total	68	1.00	100		

Observe que la suma de todas las frecuencias absolutas es igual al total de observaciones, y la suma de todas las frecuencias relativas es igual a 1 o a 100 en dependencia de si se trabaja con proporciones o porcentajes.

También pueden calcularse las frecuencias acumuladas absolutas o relativas. Para ello se suman las frecuencias de todas las clases que anteceden a la clase para la cual se está realizando el cálculo. La frecuencia acumulada del primer intervalo de clase coincide con la absoluta o relativa del mismo, pues como es obvio nadie la antecede y la del último intervalo es el total de observaciones 1 o 100.

Resumiendo, las distribuciones de frecuencias pueden clasificarse en: Distribución de frecuencia:

- Absoluta
- Relativa
- Acumulada-relativa-absoluta

Tabla de frecuencia de datos no agrupados

Los datos no agrupados son las de observaciones realizadas en un estudio estadístico que se presentan en su forma original tal y como fueron recolectados, para obtener información directamente de ellos.

La Tabla de frecuencia de datos no agrupados indica las frecuencias con que aparecen los datos estadísticos sin que se haya hecho ninguna modificación al tamaño de las unidades originales. En estas distribuciones cada dato mantiene su propia identidad después que la distribución de frecuencia se ha elaborado. En estas distribuciones los valores de cada variable han sido solamente reagrupados, siguiendo un orden lógico con sus respectivas frecuencias.

La tabla de frecuencias de datos no agrupados se emplea si las variables toman un número pequeños de valores o la variable es discreta.

Tabla de frecuencia de datos agrupados

La Tabla de frecuencia de datos agrupados aquella distribución en la que los datos estadísticos se encuentran ordenados en clases y con la frecuencia de cada clase; es decir, los datos originales de varios valores adyacentes del conjunto se combinan para formar un intervalo de clase.

“Ejemplo de Tabla de Frecuencia de datos agrupados”

La tabla de frecuencias agrupadas se emplea generalmente si las variables toman un número grande de valores o la variable es continua.

En este caso se agrupan los valores en intervalos que tengan la misma amplitud denominados clases. A cada clase se le asigna su frecuencia correspondiente.

Las clases deben ser excluyentes y exhaustivas, es decir que cada elemento de la muestra debe pertenecer a una sola clase y a su vez, todo elemento debe pertenecer a alguna clase.

Cada clase está delimitada por el límite inferior de la clase y el límite superior de la clase.

Los intervalos se forman teniendo presente que el límite inferior de una clase pertenece al intervalo, pero el límite superior no pertenece intervalo, se cuenta en el siguiente intervalo. No existe una regla fija de cuantos son los intervalos que se deben hacer; hay diferentes criterios, la literatura especializada recomienda considerar entre 5 y 20 intervalos. El número de intervalos se representa por la letra “K”.

El Recorrido es el límite dentro del cual están comprendidos todos los valores de la serie de datos. Es la diferencia entre el valor máximo de una variable y el valor mínimo que ésta toma en una investigación cualquiera.

$$R = X_{max.} - X_{min.}$$

La Amplitud de la clase es la diferencia entre el límite superior e inferior de la clase y se representarán por “ C_i ”

$$C_i = R/K$$

Se considerará la misma amplitud para todos los intervalos

La Marcas de clases (X_i) representa a la variable a través de un valor. Se calcula como el punto medio de cada clase, o bien la semi suma de la clase.

La tabla de frecuencias puede representarse gráficamente en un histograma. Normalmente en el eje vertical se coloca las frecuencias y en el horizontal los intervalos de valores.

Tablas de frecuencias para variables cualitativa

En el caso de variable cualitativa no se pueden calcular las frecuencias acumuladas pues no es posible establecer un orden en las clases dentro de la modalidad. Colocamos en la tabla aquellos valores que son independientes del lugar en que se pongan las modalidades.

Construcción de tabla de frecuencia: Interpretación, gráfica y aplicaciones. Recordemos que las variables cuantitativas o métricas pueden ser de dos tipos: continuas o discretas. En el primer caso la construcción de una tabla de distribución de frecuencias requiere de la aplicación de un proceso simple y de la definición de algunos conceptos. En el segundo caso este proceso es aún más sencillo.

Cuando la variable es continua, la construcción de una tabla de frecuencia presenta como su punto de mayor importancia la determinación del número de intervalos o clases que la formarán. Una clase o intervalo de clase es el elemento en la tabla que permite condensar en mayor grado un conjunto de datos con el propósito de hacer un resumen de ellos. El número de casos o mediciones que quedan dentro de un intervalo reciben el nombre de frecuencia del intervalo, y se denota generalmente como f_i . La diferencia entre el extremo mayor y el extremo menor del intervalo se llama longitud o anchura del intervalo. Para construir una tabla de distribución de frecuencias para una variable continua es conveniente utilizar los siguientes pasos:

- Se determina primeramente el número de intervalos o clases en la tabla en función al número de datos a condensar, para ello podemos utilizar dos criterios de selección: El primero consiste en que el investigador selecciona el número de intervalos utilizando la siguiente tabla:

Numero de datos	Numero de intervalo
De 10 a 100	De 4 a 8
De 100 a 1000	De 8 a 11
De 1000 a 10 000	De 11 a 20

El segundo, consiste en calcular la fórmula de Sturges, que determina un número

mero aproximado de intervalos k . Aunque ésta no siempre resulta muy adecuada es una relación muy utilizada. $K = 1 + 3.322 \log(n)$

Dónde: n es el número de datos a condensar en la tabla

- Una vez seleccionado el número de intervalos k , se procede a determinar su longitud o ancho. Observe que esta longitud o ancho es la misma para todos los intervalos en la tabla de frecuencia. Esto último se hace con la finalidad de facilitar los cálculos mediante métodos simplificados.

Para calcular el ancho del intervalo se aplica la siguiente relación:

$$t = \frac{\text{dato mayor} - \text{dato menor}}{k}$$

Si el valor de t , no es entero, el investigador puede manejarlo usando la fracción, o bien, seleccionar el número par más cercano a este cociente. Los intervalos no deben ser muy grandes, al grado que enmascare la distribución, ni tan pequeños que casi no contribuyan a facilitar los cálculos.

La diferencia entre el dato mayor y el menor del conjunto de datos que se analiza recibe el nombre de amplitud. Cuando el investigador desee utilizar intervalos con anchos desiguales deberá tomar en consideración que los cálculos serán más laboriosos por no ser aplicables los métodos simplificados.

- Una vez determinado el número de intervalos y su tamaño, el paso siguiente consiste en indicar el límite inferior de la primera clase, el cual puede ser un valor igual o ligeramente menor al dato de valor mínimo del conjunto de datos. Una vez hecho esto, le sumamos el valor del ancho del intervalo para fijar el límite superior de esta clase considerando en ello los valores de los límites.

Indicamos el límite inferior de la segunda clase agregando una unidad al límite superior de la primera clase. El límite superior de esta segunda clase será la suma del ancho del intervalo al límite inferior de la misma. Bajo esta dinámica se construyen todos los intervalos de clase, tomando en consideración que el

primer intervalo debe contener al menor de los datos y el último el mayor.

Los intervalos hasta aquí contruidos reciben el nombre de intervalos de clase o intervalos ficticios.

- Se construyen los intervalos reales de clase, restando media unidad a los límites inferiores de los intervalos ficticios y agregando media unidad a los límites superiores de los mismos.
- Con el establecimiento de los límites reales de clase en la tabla se efectúa la clasificación de los datos en cada intervalo para determinar así la frecuencia de cada clase (f_i).
- Finalmente se construye la tabla de frecuencias definitiva, la cual contiene, en la primer columna, la clase; en la segunda columna, los intervalos reales de clase y, en la tercer columna, las frecuencias de clase, también llamadas frecuencias absolutas.

Como complemento el investigador por medio de las tablas de distribución de frecuencias existe la posibilidad de construir gráficas de diversos tipos que le permiten explicar más fácilmente el comportamiento de los datos estudiados. Una gráfica permite mostrar, explicar, interpretar y analizar de manera sencilla, clara y efectiva los datos estadísticos mediante formas geométricas tales como líneas, áreas, volúmenes, superficies, etcétera. Las gráficas permiten además la comparación de magnitudes, tendencias y relaciones entre los valores que adquiere una variable.

Las gráficas tienen gran utilidad como medios de divulgación del análisis estadístico, ya que las relaciones visuales se captan con facilidad y resulta sencillo recordarlas. Histogramas y polígonos de frecuencias.

Gráficos para representar variables en escala cualitativa y cuantitativa discreta.

- Gráfico de barras simples: Es un gráfico formado por barras separadas que representan a las categorías de la variable en estudio. Se utiliza cuando queremos representar una variable cualitativa o cuantitativa dis-

creta, y la información se dispone en frecuencias absolutas o relativas, o en medidas de resumen.

Elementos a considerar en su construcción

1. Dispón las barras separadas entre sí, para dar la idea de discontinuidad de la variable representada.
 2. El ancho de las barras será opcional, pero debe ser el mismo para todas.
 3. La separación entre barras debe ser igual a la mitad del ancho de ellas.
 4. Si la variable es nominal, ordena las barras en orden creciente o decreciente, en dependencia de tus gustos.
 5. Utiliza tantas barras como categorías tenga la variable.
 6. Puedes colocar las barras en el eje vertical o en el horizontal. Comúnmente se utiliza el eje horizontal.
 7. Este gráfico se origina a partir de tablas unidimensionales.
- Gráfico de pastel, de sectores o circular: Este gráfico se utiliza cuando queremos representar una variable cualitativa o cuantitativa discreta, y la información se dispone en porcentaje. Básicamente, es un círculo dividido en sectores que representan las categorías de la variable.

Elementos a considerar en su construcción

1. La totalidad de la información se representa por el número total de grados de un círculo (360s).
 2. Para obtener los grados correspondientes a cada categoría, se multiplica 3.60 por la frecuencia relativa utilizada.
- Gráfico de barras múltiples: Este gráfico se utiliza cuando queremos representar dos variables, las cuales pueden ser: cualitativas o cuantitativas discretas ambas, o una cualitativa y la otra cuantitativa discreta; y la información se dispone en frecuencias absolutas o relativas, o en medidas de resumen. Los datos se representan mediante barras agrupadas, como verás a continuación.

Elementos a considerar en su construcción

1. Dispondrás grupos de dos, tres o más barras, es decir, barras dobles, triples, etc.
 2. El número de grupos a formar dependerá del número de categorías consignadas en la columna matriz o en la fila de encabezamiento, según tu gusto.
 3. La separación entre cada grupo de barras es aproximadamente la mitad del ancho del grupo.
 4. Este gráfico se origina a partir de tablas bidimensionales.
- Gráfico de barras compuestas: Al igual que el gráfico anterior, utiliza este cuando quieras representar dos variables: ambas cualitativas o cuantitativas discretas, o una cualitativa y la otra cuantitativa discreta; y dispongas la información en frecuencias relativas. Aquí, la información perteneciente a una variable se representa en su totalidad en una sola barra.

Elementos a considerar en su construcción

1. Cada barra representa el ciento por ciento de la información del grupo representado.
 2. El ancho de las barras queda a tu gusto, pero debe ser el mismo para todas.
 3. La separación entre las barras es aproximadamente la mitad del ancho.
 4. Lo originan tablas bidimensionales.
- Polígono de frecuencias: Este gráfico se utiliza cuando queremos representar hasta dos variables, de las que al menos una debe ser cuantitativa continua y la otra (en caso de ser dos variables) cualitativa o cuantitativa discreta, y la información se dispone en frecuencias absolutas o relativas, o en medidas de resumen. Está formado por una o dos curvas que representan a cada variable estudiada.

Elementos a considerar en su construcción

1. Se pueden construir histogramas inicialmente, y luego marcar los puntos medios de cada intervalo de clase, los cuales al unirse forman una curva.
2. Habrá tantas curvas como categorías tenga la variable discontinua.
3. Lo originan tablas uni o bidimensionales.

Ojivas: Polígono de frecuencia acumulada, en las abscisas se colocan los límites superiores de cada intervalo de clase y en las ordenadas se coloca la frecuencia acumulada (absoluta o relativa) de la clase.

Su utilidad radica:

- Calcular el número o porcentaje de observaciones que corresponden a un intervalo determinado de la variable.
- Calcular lo perceptibles de la distribución de los datos.

Características de las Ojivas

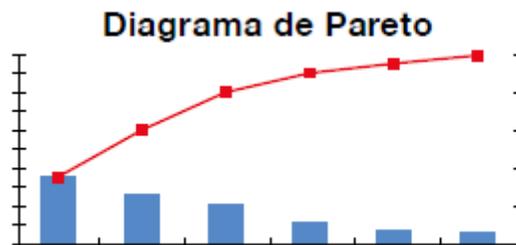
- Muestran las frecuencias acumuladas
- Se prefiere para el tratamiento de datos cuantitativos
- El punto de inicio equivale a una frecuencia de 0
- El punto final equivale al 100% de los datos.

Se convierten en una herramienta vital para el análisis estadístico.

Tabla de distribución de frecuencia interválica para variable continua.

Para una variable continua es necesario definir primero una cantidad de clases o intervalos de clase. El número de clases puede determinarse con la expresión: $K = N^{\circ} \text{Clases} \gg 1 + 3.3 \log n$. La amplitud de cada intervalo de clase se determina como $\text{Amplitud} = (R + 1) / K$, donde R es el rango o diferencia entre el valor máximo y mínimo de la variable. Otras veces el número de clases es fijado arbitrariamente, aunque siempre se pide que no sean ni muy pocas ni demasiadas.

Gráfico de Pareto y aplicación El Diagrama de Pareto es una gráfica en donde se organizan diversas clasificaciones de datos por orden descendente, de izquierda a derecha por medio de barras sencillas después de haber reunido los datos para calificar las causas. De modo que se pueda asignar un orden de prioridades.



Por lo tanto, el Análisis de Pareto es una técnica que separa los “pocos vitales” de los “muchos triviales”. Una gráfica de Pareto es utilizada para separar gráficamente los aspectos significativos de un problema desde los triviales de manera que un equipo sepa dónde dirigir sus esfuerzos para mejorar. Reducir los problemas más significativos (las barras más largas en una Gráfica Pareto) servirá más para una mejora general que reducir los más pequeños. Con frecuencia, un aspecto tendrá el 80% de los problemas. En el resto de los casos, entre 2 y 3 aspectos serán responsables por el 80% de los problemas.

¿Cuándo se utiliza?

- Al identificar un producto o servicio para el análisis para mejorar la calidad.
- Cuando existe la necesidad de llamar la atención a los problema o causas de una forma sistemática.
- Al identificar oportunidades para mejorar
- Al analizar las diferentes agrupaciones de datos (ej: por producto, por segmento, del mercado, área geográfica, etc.)
- Al buscar las causas principales de los problemas y establecer la prioridad de las soluciones
- Al evaluar los resultados de los campos efectuados a un proceso (antes

- y después) • Cuando los datos puedan clasificarse en categorías
- Cuando el rango de cada categoría es importante.

Pareto es una herramienta de análisis de datos ampliamente utilizada y es por lo tanto útil en la determinación de la causa principal durante un esfuerzo de resolución de problemas. Este permite ver cuáles son los problemas más grandes, permitiéndoles a los grupos establecer prioridades. En casos típicos, los pocos (pasos, servicios, ítems, problemas, causas) son responsables por la mayor parte del impacto negativo sobre la calidad. Si enfocamos nuestra atención en estos pocos vitales, podemos obtener la mayor ganancia potencial de nuestros esfuerzos por mejorar la calidad. Un equipo puede utilizar la Gráfica de Pareto para varios propósitos durante un proyecto para lograr mejoras:

- Para analizar las causas
- Para estudiar los resultados
- Para planear una mejora continua
- Las Gráficas de Pareto son especialmente valiosas como fotos de “antes y después” para demostrar qué progreso se ha logrado. Como tal, la Gráfica de Pareto es una herramienta sencilla pero poderosa.

¿Cómo se utiliza?

1. Seleccionar categorías lógicas para el tópico de análisis identificado (incluir el periodo de tiempo).
2. Reunir datos. La utilización de un CheckList puede ser de mucha ayuda en este paso.
3. Ordenar los datos de la mayor categoría a la menor
4. Totalizar los datos para todas las categorías
5. Calcular el porcentaje del total que cada categoría representa
6. Trazar los ejes horizontales (x) y verticales (y primario – y secundario)
7. Trazar la escala del eje vertical izquierdo para frecuencia (de 0 al total, según se calculó anteriormente)
8. De izquierda a derecha trazar las barras para cada categoría en orden descendente. Si existe una categoría “otros”, debe ser colocada al final, sin importar su valor. Es decir, que no debe tenerse en cuenta al momen-

- to de ordenar de mayor a menor la frecuencia de las categorías.
9. Trazar la escala del eje vertical derecho para el porcentaje acumulativo, comenzando por el 0 y hasta el 100%
 10. Trazar el gráfico lineal para el porcentaje acumulado, comenzando en la parte superior de la barra de la primera categoría (la mas alta)
 11. Dar un título al gráfico, agregar las fechas de cuando los datos fueron reunidos y citar la fuente de los datos.
 12. Analizar la gráfica para determinar los “pocos vitales”

Consejos para su construcción / interpretación Diagrama de Pareto es un gráfico de barras que enumera las categorías en orden descendente de izquierda a derecha, el cual puede ser utilizado por un equipo para analizar causas, estudiar resultados y planear una mejora continúa.

Dentro de las dificultades que se pueden presentar al tratar de interpretar el Diagrama de Pareto es que algunas veces los datos no indican una clara distinción entre las categorías. Esto puede verse en el gráfico cuando todas las barras son más o menos de la misma altura.

Otra dificultad es que se necesita más de la mitad de las categorías para sumar más del 60% del efecto de calidad, por lo que un buen análisis e interpretación depende en su gran mayoría de un buen análisis previo de las causas y posterior recogida de datos.

En cualquiera de los casos, parece que el principio de Pareto no aplica. Debido a que el mismo se ha demostrado como válido en literalmente miles de situaciones, es muy poco probable que se haya encontrado una excepción. Es mucho más probable que simplemente no se haya seleccionado un desglose apropiado de las categorías. Se deberá tratar de estratificar los datos de una manera diferente y repetir el Análisis de Pareto.

Esto nos lleva a la conclusión que para llevar a cabo un proceso de Resolución de Problemas /Toma de Decisiones (RP/TD) es necesario manejar cada una de las herramientas básicas de la calidad, tanto desde el punto de vista teórico como desde su aplicación.

La interpretación de un Diagrama de Pareto se puede definir completando las siguientes oraciones de ejemplo:

“Existen (número de categorías) contribuyentes relacionados con (efecto). Pero estos (número de pocos vitales) corresponden al (número) % del total (efecto). Debemos procurar estas (número) categorías pocos vitales, ya que representan la mayor ganancia potencial para nuestros esfuerzos.”

Relación con otras herramientas

Un Diagrama de Pareto generalmente se relaciona con:

- Diagrama de Causa y Efecto (Ishikawa)
- CheckList de Revisión
- CheckList de reunión de datos
- Matriz para la Planeación de Acciones

Ejemplo de aplicación

Un fabricante de heladeras desea analizar cuáles son los defectos más frecuentes que aparecen en las unidades al salir de la línea de producción. Para esto, empezó por clasificar todos los defectos posibles en sus diversos tipos:

Tipo de Defecto	Detalle del Problema
Motor no detiene	No para el motor cuando alcanza Temperatura
No enfría	El motor arranca pero la heladera no enfría
Burlete Def.	Burlete roto o deforme que no ajusta
Pintura Def.	Defectos de pintura en superficies externas
Rayas	Rayas en las superficies externas
No funciona	Al enchufar no arranca el motor
Puerta no cierra	La puerta no cierra correctamente
Gavetas Def.	Gavetas interiores con rajaduras
Motor no arranca	El motor no arranca después de ciclo de parada
Mala Nivelación	La heladera se balancea y no se puede nivelar
Puerta Def.	Puerta de refrigerador no cierra herméticamente
Otros	Otros Defectos no incluidos en los anteriores

Posteriormente, un inspector revisa cada heladera a medida que sale de producción registrando sus defectos de acuerdo con dichos tipos. Después de inspeccionar 88 heladeras, se obtuvo una tabla como esta:

Tipo de Defecto	Detalle del Problema	Frec.
Burlete Def.	Burlete roto o deforme que no ajusta	9
Pintura Def.	Defectos de pintura en superficies externas	5
Gavetas Def.	Gavetas interiores con rajaduras	1
Mala Nivelación	La heladera se balancea y no se puede nivelar	1
Motor no arranca	El motor no arranca después de ciclo de parada	1
Motor no detiene	No para el motor cuando alcanza Temperatura	36
No enfría	El motor arranca pero la heladera no enfría	27
No funciona	Al enchufar no arranca el motor	2
Otros	Otros Defectos no incluidos en los anteriores	0
Puerta Def.	Puerta de refrigerador no cierra herméticamente	0
Puerta no cierra	La puerta no cierra correctamente	2
Rayas	Rayas en las superficies externas	4
Total:		88

La última columna muestra el número de heladeras que presentaban cada tipo de defecto, es decir, la frecuencia con que se presenta cada defecto. En lugar de la frecuencia numérica podemos utilizar la frecuencia porcentual, es decir, el porcentaje de heladeras en cada tipo de defecto:

Tipo de Defecto	Detalle del Problema	Frec.	Frec. %
Burlete Def.	Burlete roto o deforme que no ajusta	9	10.2
Pintura Def.	Defectos de pintura en superficies externas	5	5.7
Gavetas Def.	Gavetas interiores con rajaduras	1	1.1
Mala Nivelación	La heladera se balancea y no se puede nivelar	1	1.1
Motor no arranca	El motor no arranca después de ciclo de parada	1	1.1
Motor no detiene	No para el motor cuando alcanza Temperatura	36	40.9
No enfría	El motor arranca pero la heladera no enfría	27	30.7
No funciona	Al enchufar no arranca el motor	2	2.3
Otros	Otros Defectos no incluidos en los anteriores	0	0.0
Puerta Def.	Puerta de refrigerador no cierra herméticamente	0	0.0
Puerta no cierra	La puerta no cierra correctamente	2	2.3
Rayas	Rayas en las superficies externas	4	4.5
Total:		88	100

Pero ¿Cuáles son los defectos que aparecen con mayor frecuencia? Para hacerlo más evidente, antes de graficar podemos ordenar los datos de la tabla en orden decreciente de frecuencia:

Vemos que la categoría “otros” siempre debe ir al final, sin importar su valor. De esta manera, si hubiese tenido un valor más alto, igual debería haberse ubicado en la última fila.

Ahora resulta evidente cuales son los tipos de defectos más frecuentes. Podemos observar que los 3 primeros tipos de defectos se presentan en el 82 % de las heladeras, aproximadamente. Por el Principio de Pareto, concluimos que: La mayor parte de los defectos encontrados en el lote pertenece sólo a 3 tipos de defectos, de manera que si se eliminan las causas que los provocan desaparecería la mayor parte de los defectos.

BIBLIOGRAFÍA

DEFINICIONES ESTADÍSTICAS Y

PRESENTACIÓN DE DATOS ESTADÍSTICOS



www.mawil.us

1. Concepto de estadística. Introducción Cap. I [citado 23 enero 2019]. Disponible en :<https://tarwi.lamolina.edu.pe/~fmendiburu/indexfiler/academic/Estadistica/parte1.pdf>
2. Samperio Pacheco V, Pérez Olguín N , Zúñiga Mera A, Hernández Serrano M , Domínguez Narváez J . Introducción a la estadística . 2011. [citado 23 enero 2019]. Disponible en :https://www.uaeh.edu.mx/docencia/P_Presentaciones/tlahuelilpan/sistemas/probabilidad_estadistica/introduccion_estadistica.pdf
3. Laguna C .Introducción a la estadística . 2014 [citado 2 Mar 2019]. Disponible en : <http://www.ics-aragon.com/cursos/salud-publica/2014/pdf/M2T01.pdf>
4. Berger, James O., , “Bayesian analysis: a look a today and thoughts of tomorrow”, en AdrianE., RAFTERY, Martin A. TANNER y Martin T. WELLS (eds.),Statistics in the 21st Century, Chapman and Press Hall, Londres. 2002
5. Anderson D , Sweeney D , Williams T . Estadística para la administración y la economía.10 a edición 2015 [citado 2 Mar 2019]. Disponible en :<https://www.upg.mx/wp-content/uploads/2015/10/LIBRO-13-Estadistica-para-administracion-y-economia.pdf>
6. Barreto-Villanueva, Adán. El progreso de la Estadística y su utilidad en la evaluación del desarrollo. Papeles de población. 2012[citado 2 Mar 2019]; 18(73). Disponible en : 241-271. http://www.scielo.org.mx/scielo.php?script=sci_arttext&pid=S1405-74252012000300010&lng=es&tlng=e
7. GlejbermanD . La investigación estadística. [citado 2 Mar 2019]. Disponible en: :http://www.dgeec.gov.py/convocatoria/document/etapas_invest_estadistica.pdf
8. BayarreVea H , Oliva Pérez M , HorsfordSaing R, Ranero Aparicio V , Coutin Marie G , D[az Llanes G. Metodología de Investigación en APS . 2004.
9. Walpole, R. E. & Myers, R. H. Probabilidad y Estadística. 4. ed. Ciudad de México, McGraw- Hill, 1996.
10. Bai, X.; Tsiatis, A. A. & O’Brien, S. M. Doubly-robust estimators of treatment-specific survival distributions in observational studies with stratified sampling. Biometrics, 69(4):830-9, 2013.

11. Ávila Baray, H. L. Introducción a la Metodología de la Investigación. Edición electrónica. Cuauhtémoc (Chihuahua), Instituto Tecnológico de Cd. Cuauhtémoc, 2006 [citado 2 Mar 2019]. Disponible en: <http://www.eumed.net/libros-gratis/2006c/203/index.htm>
12. Hund, L.; Bedrick, E. J. & Pagano, M. Choosing a cluster sampling design for lot quality assurance sampling surveys. *PLoS One*, 10(6):e0129564, 2015.
13. Canal Díaz N . Técnicas de muestreo . Sesgos más frecuentes .2006 [citado 2 Mar 2019]. Disponible en :<http://www.revistaseden.org/files/9-cap%209.pdf>
14. Otzen Tamara, Manterola Carlos. Técnicas de Muestreo sobre una Población a Estudio. *Int. J. Morphol.* [Internet]. 2017 Mar [citado 2019 Ene 31] ; 35(1): 227-232. Disponible en: https://scielo.conicyt.cl/scielo.php?script=sci_arttext&pid=S0717-95022017000100037&lng=es. <http://dx.doi.org/10.4067/S0717-95022017000100037>.
15. Gallo Pimentel J, González Díaz E . Introducción a la metodología de investigación pedagógica y técnica. 2000. Ciudad de la Habana
16. Otzen Tamara, Manterola Carlos. Técnicas de Muestreo sobre una Población a Estudio. *Int. J. Morphol.* .2017 Mar [citado 2019 Ene 31] ; 35(1): 227-232. Disponible en: https://scielo.conicyt.cl/scielo.php?script=sci_arttext&pid=S0717-95022017000100037&lng=es. <http://dx.doi.org/10.4067/S0717-95022017000100037>
17. Azcona, Maximiliano; Manzini, Fernando y Dorati, Javier. Precisiones metodológicas sobre la unidad de análisis y la unidad de observación. Aplicación a la investigación en psicología. 2013 [citado 2 Mar 2019]. Disponible en : http://sedici.unlp.edu.ar/bitstream/handle/10915/45512/Documento_completo.pdf?sequence=1
18. Presentación de datos estadísticos en cuadros y gráficas / Instituto Nacional de Estadística y Geografía México: INEGI, 2011 [citado 2 Mar 2019]. Disponible en :https://www.snieg.mx/contenidos/espanol/normatividad/doctos_genbasica/cuadros_graficas.pdf
19. Representación gráfica de datos estadísticos. 2015 [citado 2 Mar 2019]. Disponible en:<https://www.hiru.eus/es/matematicas/representacion-grafica-de-datos-estadisticos>
20. Guía para la presentación de información estadística. Segunda edición.

- 2017 [citado 2 Mar 2019]. Disponible en:<http://www.inec.go.cr/sites/default/files/documetos-bibliotecavirtual/mepresentinfoestadist-21122017.pdf>
21. Hernández Samperi R . Metodología de la Investigación. 1998. ISBN 970-10 -1899-0
 22. Artilles Visbal L. Otero Iglesias J , Barrios Osuna I . Metodología de la Investigación. Editorial Ciencias Médicas , la habana 2015. ISBN 978-959-212-385-4
 23. Fraga Rodríguez R, Herrera Padron M , Metodología de la investigación educativa. 1999
 24. Técnicas de registros.[citado 2 Mar 2019]. Disponible en :<https://www.google.com/search?q=metod+de+recoleccion+de+datos.+Registro+continuo++&ie=utf-8&oe=utf-8>
 25. Ana M. Bianco y Elena J. Martínez. Etapas de una investigación. 2004 [citado 2 Mar 2019]. Disponible en :http://www.dm.uba.ar/materias/probabilidades_estadistica_C/2011/1/PyEC12.pdf
 26. González Castellanos M , Yll Lavín M , Curiel Lorenzo L . Metodología e la investigación científica para las ciencias técnicas . 2003 [citado 2 Mar 2019]. Disponible en : [http:// www.sld.cu/galerias/pdf/sitios/bmn/metodologia_de_la_investigacion.diseno_teorico_y_formulacion_proyecto_investigacion.pdf](http://www.sld.cu/galerias/pdf/sitios/bmn/metodologia_de_la_investigacion.diseno_teorico_y_formulacion_proyecto_investigacion.pdf)
 27. Cáceres Ureña, F . Las fuentes. Censo de población .II Reunión regional sobre evaluación y estimaciones demográficas con base en información censal IX Censo Nacional de. 2013. [citado 2 Mar 2019]. Disponible en : https://www.cepal.org/celade/noticias/paginas/5/51425/Francisco_CaceresS10RepDom.pdf
 28. Castillo S . TABLAS DE DISTRIBUCIÓN Y REPRESENTACIONES GRÁFICAS. 2011. [citado 2 Mar 2019]. Disponible en :<http://mat120estadisticaaplicada.blogspot.com/2011/04/2-representaciones-graficas.html>
 29. Distribución de frecuencias y gráficas en estadística. [citado 2 Mar 2019]. Disponible en:<http://univirtual.utp.edu.co/pandora/recursos/1000/1792/1792.pdf>
 30. Ojiva con Frecuencia Acumulada [citado 2 Mar 2019]. Disponible en : <https://sites.google.com/site/graficasdeyasodhra/tipo-de-graficas/oji->

va-con-frecuencia-acumulada

31. Análisis Exploratorio de Datos – Estadística Descriptiva. [citado 2 Mar 2019]. Disponible en: <http://www.fca.proed.unc.edu.ar/mod/book/view.php?id=3270&chapterid=205>
32. Sales M . Diagrama de Pareto . 2002 [citado 2 Mar 2019]. Disponible en : <https://www.gestiopolis.com/diagrama-de-pareto/>
33. Bonet Borjas, Carlos Manuel Ley de Pareto aplicada a la fiabilidad Ingeniería Mecánica, vol.8, núm. 3, septiembre-diciembre, 2005, pp. 1-19 Instituto Superior Politécnico José Antonio Echeverría Ciudad de La Habana, Cuba

UNIDAD II

INDICADORES DE TENDENCIA CENTRAL Y DE VARIABILIDAD



www.mawil.us

Indicadores de tendencia central

Están constituidas por un número al que sea cercano "tienden" la mayoría de las observaciones. Con otras palabras, alrededor de él se agrupan las observaciones de la serie de datos, puesto que es en si el centro de la serie, aunque ello no significa que este número tiene que estar representado en la serie(dehecho, muchas veces no ocurre así).



Las medidas de tendencia central nos indican, en forma clara y concisa, la ubicación o posición del valor central o el punto de gravedad de la serie estadística. Cada una tiene sus ventajas y limitaciones y al utilizarlas debemos considerar, en primera instancia, la escala de medición de los datos.

Veamos en detalle cada una de estas medidas

Moda

Es el valor o característica más frecuente de la serie estadística. Esta definición es válida tanto para series cuantitativas como para series cualitativas. La moda tiene escasa o ninguna utilidad desde el punto de vista matemático o de prueba de hipótesis sobre la distribución que caracteriza a los datos

Entre las ventajas de la moda tenemos que es fácil de calcular y que es la única medida de posición central apropiada para describir atributos (series cua-

litativas). No es posible determinar ni la mediana ni la media para atributos tales como sexo, especie, tipo de bosque o color de las hojas

$$m_o = l_i + ((f_i - f_{i-1}) / (f_i - f_{i-1} + f_i - f_{i+1})) * t_i$$

l_i Extremo inferior del intervalo modal (intervalo que tiene mayor frecuencia absoluta).

f_i Frecuencia absoluta del intervalo modal.

f_{i-1} Frecuencia absoluta del intervalo anterior al modal.

f_{i+1} Frecuencia absoluta del intervalo posterior al modal.

t_i Amplitud de los intervalos.

La media aritmética

Se puede encontrar con diversos nombres, entre ellos los más utilizados son promedio, promedio aritmético e incluso simplemente media.

La media aritmética es una cifra que se obtiene al sumar todos los valores observados y dividirlos por el número de valores. Se denota por los símbolos μ (letra griega mu) X (se lee equismedia), el primero se utiliza cuando trabajamos con toda la población objeto de estudio y el segundo cuando trabajamos con una muestra de la población, que es lo más frecuente. La media conserva las unidades de medida de la variable en su estado original, o sea, que la media de un grupo de edad ese años se expresare asimismo en años.

La fórmula del cálculo de la media es:

Media (\bar{x}) = $\sum X * f / N$ Siendo X las observaciones distintas f las frecuencias relativas y N el número total de datos. La media ponderada es una variación de la media aritmética simple y se utiliza cuando las observaciones no tienen el mismo peso o importancia. Su fórmula es:

$$\bar{X} = \sum (W_i * X_i)$$

Dónde: W_i es el peso o ponderación dado a la observación X_i La media aritmética para datos agrupados es un tipo de media ponderada en donde el factor

de ponderación es la frecuencia de cada clase. Matemáticamente se expresa así:

$$\bar{X} = \Sigma (f_i * m_i)/n$$

En donde: m_i representa el punto medio de la clase i y f_i su frecuencia. La media aritmética es el punto de balance de la distribución; matemáticamente esto significa que la desviación de los valores a su alrededor es siempre cero. Esto es, si calculamos la desviación de cada observación con respecto a la media, la suma de las desviaciones será igual al cero. Por esta razón si cambiamos uno de los valores que componen la serie estadística, la media también cambiará.

Entre las propiedades de la media tenemos las siguientes:

1. Es fácilmente comprensible por la mayoría de las personas (o, al menos, es fácil de explicar su significado).
2. Siempre existe, y puede calcularse para cualquier grupo de datos numéricos.
3. Es única, o sea, un grupo de datos solo tiene una media;
4. Toma en cuenta a todos los valores de la serie de forma individual, esto es, recorre la serie completa.

Resulta sumamente importante, el cálculo de la media ya que representa a todos los valores de la serie, siendo precisamente lo que se quiere lograr. Ahora bien, no siempre esto resulta beneficioso, si existen valores muy alejados de la mayoría (datos aberrantes), entonces se distorsiona mucho y deja de reflejar la realidad existente, por lo que debe emplearse otro tipo de medida de tendencia central, la mediana.

Mediana

La mediana (M_d) es el valor que divide la serie estadística ordenada en dos partes iguales, o sea, 50% de las observaciones son mayores y 50% menores que la mediana, es la observación que ocupa la posición central de una serie ordenada.

Al igual que la media, se utiliza para describir el “centro” de un grupo de datos.

Para el cálculo de la misma es ordenar la serie, ya sea en orden creciente o decreciente. Luego, buscar cuál de los valores es la mediana, lo cual dependerá del número total de observaciones o datos que tengas.

$$Md = (n+1)/2$$

Para los siguientes valores de diámetro (cm), la mediana es la observación número tres: $(5+1)/2=3$, cuyo valor es 8 cm:

5, 6, 8, 10, 11

Cuando el número de observaciones es par, como por ejemplo 10, 20 o 40, la mediana corresponde al punto medio de los valores medios de la serie por ejemplo: $(10+1)/2 = 5,5$. Para la siguiente serie estadística la mediana es el valor de la observación 5,5 o sea: $(20+21)/2 = 20,5$

10, 18, 19, 19, 20, 21, 22, 28, 33, 34.

Mediana

Para datos agrupados la mediana se calcula utilizando la curva de frecuencia acumulada o la siguiente expresión:

$$Md = Li + \{Ic \times [(0.5 \wedge n) _ \text{facum.inf}]\} / \text{fmd}$$

Donde:

Li = límite real inferior de la clase que contiene a la mediana (50% de las observaciones).

Ic= intervalo de la clase mediana

n = número de observaciones.

fmd = frecuencia absoluta de la clase mediana

$facum.inf$ = frecuencia acumulada por debajo de la clase mediana

Una desventaja de la mediana es su insensibilidad a valores extremos.

La mediana posee las siguientes propiedades:

1. Su cálculo es sencillo;
2. Siempre existe, y puedes calcularla a cualquier conjunto de datos numéricos;
3. Es única;
4. No se afecta fácilmente por valores extremos.

Características de la Mediana y Media

Características	Medida de posición central	
	Mediana	Media
Derivación matemática	No	Si
Estabilidad muestral	Regular	Excelente
Sensibilidad a valores extremos	Poco	Muy alta
Aplicación	Intervalo, Razón	Razón Intervalo
Nivel de inferencia	No paramétrico	Paramétrico

La derivación matemática se refiere a que el valor se obtiene de la interacción de todos los valores que componen la serie. La estabilidad muestral indica la estabilidad del estimador ante factores aleatorios (cambio del estimador entre muestras)

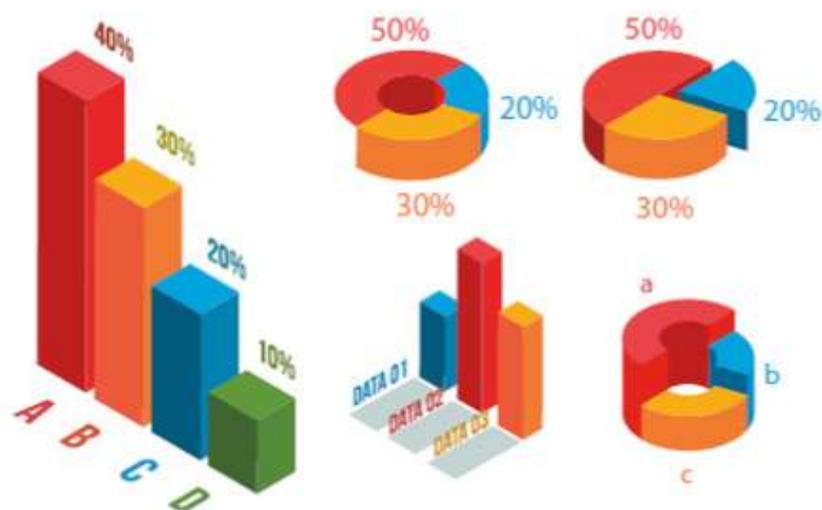
Variabilidad

Las medidas de variabilidad expresan cuantitativamente el grado de dispersión o agrupamiento de los datos. Conocer la variabilidad de un set de datos es esencial para describirlo, compararlo, realizar pruebas estadísticas y determinar tamaños de muestra.

Por ejemplo, supongamos que tenemos dos instrumentos para medir alturas, A y B, y deseamos seleccionar el más preciso (el de menor variabilidad). Al realizar una serie de mediciones en un grupo de 25 árboles con una altura media

de 15,0 metros, encontramos que la media aritmética para el instrumento A es 14,9 m y para el B 15,1 metros.

Basados en estos datos, podríamos seleccionar cualquiera de los dos instrumentos; ya que sus medias son muy similares y podríamos concluir que un instrumento es tan “bueno” como el otro. Sin embargo, la media no nos brinda información sobre la precisión del instrumento. Si además de la media observamos la distribución de alturas para cada instrumento (Fig.), seleccionaríamos el “A”; ya que muestra un menor grado de dispersión y por ende sería el más preciso.



Es oportuno señalar que la media por sí solo no es suficiente para caracterizar una serie estadística y que se requiere de otra medida que nos indique el grado de variabilidad de los datos.

El concepto de variabilidad es uno de los aspectos claves sobre los cuales descansa la estadística inferencial; ya que permite estimar el error estadístico asociado a las mediciones. Las medidas de variabilidad expresan el grado de dispersión de los datos con respecto a su centro de gravedad.

BIBLIOGRAFÍA

INDICADORES DE TENDENCIA CENTRAL Y DE VARIABILIDAD.



www.mawil.us

1. Fallas J. Tendencia Central, Variabilidad y forma de la distribución. 2012. [citado 2 Mar 2019]. Disponible en : http://www.ucipfg.com/Repositorio/MGAP/MGAP-05/BLOQUE-ACADEMICO/Unidad-2/complementarias/tendencia_central_y_variabilidad_mayo28_2012.pdf
2. Bayarre Vea H , Oliva Pérez M , Horsford Saing M , Ranero Aparicio V , Coutin Marie G , Diaz Llanes G . Tema I . Estadística Descriptiva. Metodología de la Investigación en APS. 2004
3. Artilles Visbal L. Otero Iglesias J , Barrios Osuna I . Metodología de la Investigación. Editorial Ciencias Médicas , la habana 2015. ISBN 978-959-212-385-4

UNIDAD III

NUMERO ÍNDICES Y ANÁLISIS DE REGRESIÓN SIMPLE



www.mawil.us

Numero Índices

El número índice ha sido empleado con frecuencia como un indicador bastante aproximado para estudiar la tendencia de eventos cuyo comportamiento es muy fluctuante. Para confeccionarlo se escoge el valor de algún año y este se considera entonces como el cien por ciento obteniéndose a continuación el porcentaje de éste que representa cada año de la serie. Este método tiene la ventaja de que permite apreciar la evolución del evento a lo largo del tiempo. Sin embargo, tiene una gran desventaja y ella está dada por la decisión de cuál el año que debe ser considerado como el índice. Supongamos que se escoge una no que tiene un valor muy alto los demás años estarán por debajo del mismo o viceversa, no obstante, esto no resulta tan problemático si se hace la interpretación adecuada.

Un número índice es un valor relativo, expresado como porcentaje o cociente, que mide un período dado contra un período base determinado

Uno de los problemas de mayor importancia a la hora de elaborar un número índice es el conseguir que éste sea adecuadamente representativo, para ello es preciso que el índice cumpla ciertas propiedades de carácter matemático y reúna ciertos requisitos en su definición:

- **Identidad.** Cuando el período base y el de comparación coinciden, el índice debe ser igual a uno.
- **Inversión.** Si en un índice se invierten los períodos base y de comparación, el índice toma el valor recíproco al anterior.
- **Circular.** Si se multiplica el índice de un período Z con relación a un período Y por el índice de Y con relación a X, el producto ha de ser el índice de Z con relación a X.
- **Existencia.** El índice ha de tomar valores reales y finitos para cualquier valor de la variable observada.
- **Proporcionalidad.** El índice elaborado sobre unos determinados valores de una variable ha de ser proporcional al índice correspondiente a los valores de esa variable multiplicados por un mismo número K.
- **Variación proporcional.** Si los valores de la variable varían en una cierta

cuantía, el índice varía proporcionalmente.

- Inalterabilidad. Si se introduce una nueva modalidad en el índice complejo, de tal manera que el valor de éste coincide con el del índice simple de aquella, el índice complejo no varía.
- Homogeneidad. El valor de un índice no ha de ser afectado por modificaciones de las unidades de medida.

Índice simple

Se conceptualiza como el cociente entre la magnitud en el período corriente y la magnitud en el período base. Generalmente se multiplica por cien y se lee en porcentaje. No presentan gran utilidad en sí mismos y su interés radica en que son el punto de partida de la construcción de los índices complejos y en que algunas de sus propiedades sirven para evaluar la bondad de éstos.

Al período inicial se le denomina período base o referencia y se le asigna el valor 100, en cambio, la situación que deseamos comparar se denomina período actual o corriente.

Para las comparaciones hay que tener en cuenta dos aspectos importantes:

- Fijar la situación inicial (de forma arbitraria) a la que se referirán las comparaciones. Señalar que la elección de la situación inicial condiciona el resultado de la comparación, por lo que el punto de referencia inicial debe ser el más idóneo posible a los objetivos que se persiguen.
- Las magnitudes que se comparan pueden ser simples o complejas, lo que nos introduce en el problema de la construcción de sistemas de comparación adecuados. Una magnitud compleja es comparar la producción de un mismo país en dos épocas diferentes o la producción global de dos países. No olvidemos que la producción es una magnitud compleja compuesta por magnitudes simples heterogéneas (unidades de producción, litros, kilogramos, etc.)

Índices simples más utilizados Precio relativo: Relación entre el precio de un bien en el período actual p_{it} y el precio del mismo en el período base:

$$p_{i_0}: p_0^t = p_{it} / p_{i_0} \cdot 100$$

Cantidad relativa: Razón entre la cantidad producida o vendida de un bien en sus períodos actual.

$$q_{i_0}: q_0^t = q_{it} / q_{i_0} \cdot 100$$

Valor relativo: Valor de un bien en un período cualquiera se define como el producto del precio de ese bien y la cantidad producida (vendida). El valor relativo será la razón entre los valores de ese bien en el período actual ($p_{it} \cdot q_{it}$) y en el periodo de base ($p_{i_0} \cdot q_{i_0}$).

$$V_0^t = V_i / V_0 = p_{it} \cdot q_{it} / p_{i_0} \cdot q_{i_0} \cdot 100 = (p_{it} / p_{i_0}) \cdot (q_{it} / q_{i_0}) \cdot 100 = p_0^t \cdot q_0^t \cdot 100$$

El valor relativo de un bien es igual al producto de su precio relativo y su cantidad relativa.

Ejemplo

Coste de la vida en una ciudad el presente año en comparación con la del año anterior (tiempo).

Coste de la vida en una ciudad comparado con el coste de la vida en otra ciudad (espacio).

Numero de índices simples: Precio, cantidad y Valor

En caso que la variable en consideración corresponda a un precio, entonces el índice simple se conoce como un índice (simple) de precios. Si la variable considerada son las cantidades, entonces se le denomina un índice (simple) de cantidades (o también conocido como un índice de volumen). Si por el contrario la variable bajo consideración es el valor en pesos de una transacción (precio multiplicado por cantidades) entonces el índice se denominará un índice (simple) de valor.

El Índice de Precios

El Índice de precios mide la evolución de los precios de un conjunto de bienes y servicios representativos del gasto de consumo de los hogares residentes en un determinado país, ciudad o provincia. El IP mide cómo evolucionan –en promedio– los precios de esa canasta, pero no cuánto vale en un momento del tiempo. Cuando el índice sube, refleja una disminución en el poder de compra del dinero en función de los precios de ese conjunto de bienes y servicios de consumo; cuando baja, refleja un aumento del poder de compra del dinero en esos mismos términos.

Simplemente se trata de relativizar los precios, las cantidades o los valores respecto del año base.

Ejemplificamos:

Comparar el precio de un artículo hace 10 años contra su precio actual, comparar la cantidad de materia prima consumida hace 2 años contra la cantidad consumida actualmente, o bien, comparar el poder adquisitivo del sueldo hace tres años contra el poder actual.

Números de índice de cantidad y de valor Las formulas descritas previamente para la obtención de números de índice de precios se modifican fácilmente parahallar números de índice de cantidado volumen intercambiando simplemente p y q.

$$\text{Índice de media aritmética simple: } \frac{1}{k} \sum_{i=1}^k \frac{q_{it}}{q_{i0}}$$

Análogamente, el índice de agregación ponderada de cantidad con pesos del año base es:

$$\text{Índice de volumen de Laspeyres: } \frac{\sum_i^k q_{it} p_{i0}}{\sum_i^k q_{i0} p_{i0}}$$

Y el índice de agregación ponderada de cantidades con pesos en el año dado es:

Índice de volumen de Paasche: $\Sigma_i^k = 1q_{it} p_{it} / \Sigma_i^k = 1q_{i0} p_{it}$

Exactamente igual que se hace con los números de índice de precios de cantidad, se pueden definir índices de valor. El más sencillo de ellos es:

Índice de valor: $\Sigma_i^k = 1q_{it} p_{it} / \Sigma_i^k = 1q_{i0} p_{i0}$

Es importante anotar la estrecha relación que existe entre los índices simples de precios, cantidades y valor. Es claro que el valor de una transacción en el período t corresponde a:

$$V_t = Q_t P_t$$

Ahora dividamos ambos lados de la ecuación por el valor de la transacción en el período base t_0 , es decir:

$$V_t / V_{t_0} = Q_t P_t / V_{t_0} = Q_t P_t / Q_{t_0} P_{t_0}$$

Multiplicando ambos lados de la ecuación por 100 tendremos:

$$100 \cdot V_t / V_{t_0} = 100 \cdot Q_t / Q_{t_0} \cdot P_t / P_{t_0}$$

Y, empleando la definición de los índices de precios, cantidades y de valor, tenemos que:

$$100 \cdot V_t / V_{t_0} = 100 \cdot Q_t / Q_{t_0} \cdot 100 / 100 \cdot P_t / P_{t_0} \cdot 100/100 (*)$$

Es decir, si conocemos dos de los índices, podremos encontrar el tercero por medio de la relación expresada en (*)

Índice de base móvil

Una media móvil es un indicador técnico que combina precios de un activo a lo largo de un período de tiempo establecido, y los divide entre el número de datos recogidos para dar una línea de tendencia. Es popular entre los inversores

porque puede ayudar a determinar la dirección de una tendencia actual y reducir el impacto de subidas de precio repentinas.

Analizando movimientos previos en el precio de un activo, una media móvil le permitirá examinar sus niveles de soporte y resistencia. Es una medida de cambio que rastrea el precio anterior de un activo y evalúa el historial de los movimientos de mercado para calcular posibles patrones futuros. Una media móvil es principalmente un indicador rezagado, lo que la convierte en una de las herramientas más populares de análisis técnico.

Este indicador se denomina ‘móvil’ porque conforme se van introduciendo datos más recientes en la gráfica, éstos remplazan a los más antiguos³.

Media geométrica de índice

La media geométrica tiene propiedades que la hacen especialmente útil en los estudios de economía, en sicología y en las ciencias sociales. En economía se emplea, por ejemplo, para calcular el promedio de porcentajes o el promedio de Índice de Precios al Consumidor (IPC) a lo largo de una serie de años

Número de índices compuesto: índices no ponderados e índice ponderado⁸.

El índice compuesto es, en realidad, el de mayor importancia. Entre los índices compuestos a los que se les presta mayor atención están: El estimador mensual de actividad económica; el índice de precios al por mayor; el índice de precios al consumidor.

Los índices agregados no ponderados o no pesados significan que todos los valores considerados son de igual importancia. Agregado significa que agregamos o sumamos todos los valores. La principal ventaja de este índice es su simplicidad.

Para construir un índice de precios agrega dos sin ponderar, primero debemos obtener la suma de los diversos precios para cada uno de los periodos que se consideran y luego dividirla por la suma de los precios del periodo base.

Sea p_0 la suma de los precios del periodo base y sea p_n la suma de los precios del periodo dado; el cociente de las dos sumas multiplicado por 100 arroja el índice P expresado en porcentaje; esto es:

$$P = \frac{\sum p_n}{\sum p_0} * 100$$

A partir de los datos del siguiente cuadro, se elaborara un índice sin ponderar Cuadro I.

Artículo	Unidad de medida	Abril 06	Abril 07
	P_0	P_n	
Jamón cocido	kg	21.81	21.96
Paleta	kg	8.53	8.43
Prepizza	Unidad	1.25	1.45
Filet de merlusa	kg	10.85	13.59
Total		42.44	45.46

Índice de precio sin ponderar

$$P = \frac{\sum 45.46}{\sum 42.44} * 100 = 107.8$$

Como se puede ver se trata de un índice muy simple y solamente se puede usar excepcionalmente, pues al no aplicar ponderaciones está sujeto a errores si no hay homogeneidad en la importancia de los distintos artículos y en las unidades. En síntesis, se dice que, en general, es un índice no recomendable.

Una manera de rectificar las desventajas de un índice agregado no ponderado es la de construir un promedio simple de porcentajes relativos. Para calcular tal índice, se convierten los precios reales de cada variable en porcentajes del periodo base, los cuales se llaman relativos porque se calculan respecto del valor del periodo base. Se obtiene un precio relativo por ejemplo, al dividir el precio de un artículo para un periodo dado P_n por el precio P_0 del precio base. La suma de todos los precios relativos, o sea $\sum P_n/P_0$ dividida por n , número de artículos que entran en el cálculo, y multiplicada por 100, arroja el promedio simple de precios relativos P , esto es:

$$P = \frac{\sum p_n}{\sum p_0} * 100$$

Que es la media aritmética de los porcentajes relativos. En el siguiente cuadro se realiza el cálculo de un promedio simple de precios relativos para los datos del cuadro II

Cuadro II

Artículo	Unidad de medida	Abril 06	Abril 07	Abril 06	Abril 07
	P0	Pn	(P0/p0)*100	(Pn/p0)*100	
Jamón cocido	kg	21.81	21.96	100	100.69
Paleta	kg	8.53	8.43	100	99.18
Prepizza	Unidad	1.25	1.45	100	106.00
Filet de merlusa	kg	10.85	13.59	100	125.25
Suma global		42.44	45.46	400	441.12
Índice				100	110.28

Este promedio simple de porcentajes, a pesar de tener la ventaja de que son números adimensionales tiene la desventaja de suponer que cada uno de los porcentajes relativos tiene la misma importancia. Es decir, artículos secundarios tienen la misma influencia o peso en el resultado final que un artículo de primera necesidad. De esta manera existen “ponderaciones ocultas” que son inadmisibles y que afectan la utilidad de este tipo de índices.

Índices compuestos ponderados

En el cálculo de los índices compuestos ponderados intervienen unos pesos w_i para cada variable X_i , pesos, que a su vez, puede ser constantes en el tiempo, o bien variable en cada periodo.

El principal interés de los índices ponderados es el hecho de poder resaltar o alternar la influencia de las diferentes cantidades, de acuerdo con algún criterio externo.

En el caso concreto de los índices de precios, los criterios más empleados

para ponderar son:

1. $w_i = p_{i_0} q_{i_0}$ que corresponde a dar un peso a la Variable X1 proporcional al precio de venta en el periodo base p_{i_0} multiplicando por la cantidad en el periodo base q_{i_0} . El peso es contante para la variable X1 a lo largo del tiempo y el factor de peso es el valor en el periodo base.

El Índice de Laspeyres o método del año base corresponde a elegir este criterio de ponderación sobre el índice de la media aritmética ponderada.

$$L_p = \frac{\sum_{i=1}^k I_i w_i}{\sum_{i=1}^k w_i} = \frac{\sum_{i=1}^k (p_{it} / p_{i_0}) p_{i_0} q_{i_0}}{\sum_{i=1}^k p_{i_0} q_{i_0}} = \frac{\sum_{i=1}^k p_{it} q_{i_0}}{\sum_{i=1}^k p_{i_0} q_{i_0}}$$

A la vista del resultado, el Índice de Laspeyres es también el índice de precios por agregación ponderada con los pesos de cantidad en el año base

2. $w_i = p_{i_0} q_{it}$ que corresponde a dar un peso a la Variable X1 proporcional al precio de venta en el periodo base p_{i_0} multiplicando por la cantidad vendida en el periodo actual q_{it} . El peso es variable para cada variable X1, a lo largo del tiempo y el factor de peso es el valor calculado a precio del periodo base y cantidad anual.

El índice de Paasche o método del año dado corresponde a elegir este criterio de ponderación sobre el índice de la media aritmética ponderada:

$$I_p = \frac{\sum_{i=1}^k I_i w_i}{\sum_{i=1}^k w_i} = \frac{\sum_{i=1}^k (p_{it} / p_{i_0}) p_{i_0} q_{is}}{\sum_{i=1}^k p_{i_0} q_{is}} = \frac{\sum_{i=1}^k p_{it} q_{is}}{\sum_{i=1}^k p_{i_0} q_{is}}$$

A la vista del resultado, el índice de Paasche es también el índice de precios por agregación ponderada con los pesos de cantidad en el año dado q_{it} .

3. $w_i = p_{i_0} q_{is}$ que corresponde a dar un peso a la Variable X1 proporcional al precio de venta en el periodo base p_{i_0} multiplicando por la cantidad vendida en año típico q_{is} . El peso es contante para cada variable X1 a lo largo del tiempo y el factor de peso es el valor calculado a precio del periodo base y cantidad en un año típico. Por esto, este índice se conoce como el método del año típico o índice de precios por agregación pon-

derada con los pesos de cantidad en el año típico q_{is} .

El Índice de Laspeyres requiere los datos de cantidad para un solo año y es más fácil de calcular. Por tanto, se utiliza con más frecuencia que el de Paasche. Como siempre se utilizó las cantidades del periodo base, se permite con el tiempo más comparaciones significativas.

Fortalezas y Debilidades relativas de los índices de Laspeyres y de Paasche

	Ventajas	Desventajas
Laspeyres	Requiere datos de cantidad para un solo periodo; por tanto: 1) Los datos se obtienen más fácilmente. 2) Se puede hacer una comparación más significativa debido a que los cambios se pueden atribuir a los movimientos en precios	Pondera los productos cuyos precios aumentan. No reflejan los cambios en los patrones de compra a través del tiempo
Paasche	Refleja los cambios en los hábitos de compra debido a que utiliza los datos de cantidad para cada periodo de referencia.	Requiere datos de cantidad para cada año; estos datos con frecuencia son difíciles de obtener. Debido a que se utilizan cantidades diferentes, es imposible atribuir las diferencias en el índice solo a los cambios en precio. Sobrepondera los productos cuyos precios disminuyen.

Índice ideal de Fisher

Entre tanto índice, parece lógico plantearse algún criterio para su elección. Desde un punto de vista teórico es deseable que los números índice para grupos de artículos tenga las propiedades que cumplían los números índice para un solo artículo. No se conoce ningún índice que cumpla con todos los criterios, si bien en muchos casos se satisfacen aproximadamente.

El índice ideal de Fisher para los precios se define como la media geométrica de los números índices Laspeyres y de Paasche.

Índice ideal de Fisher =

$$\sqrt{(\sum_{i=1}^k p_{it} q_{i0} / \sum_{i=1}^k p_{i0} q_{i0})(\sum_{i=1}^k p_{it} q_{it} / \sum_{i=1}^k p_{i0} q_{it})}$$

Como se expresó anteriormente, el índice Laspeyres tiende a sobreponderar los bienes cuyos precios aumenta, a que este incremento el precio va acompa-

ñado de una reducción en la cantidad, que no se ve reflejada en el índice Laspeyres que utiliza cantidades como base fija como ponderación. Por otra parte, el índice de Paasche tiende a sobre ponderar los productos cuyos precios baja. El Índice ideal de Fisher es un esfuerzo por compensar estos hechos. Sin embargo, la interpretación del Índice ideal de Fisher está sujeta a discusión. Por este motivo no se utilizar ampliamente.

Índices compuesto no ponderados

Para resumir la información obtenida a través de los índices simples de cada bien, estos índices promedian los simples, utilizando para ello la media aritmética, media Geométrica, media armónica y media agregativa

Consideremos N magnitudes simples, X_1, X_2, \dots, X_n , con valores $x_{10}, x_{20}, \dots, x_{n0}$, en el periodo base, y valores $x_{1t}, x_{2t}, \dots, x_{nt}$, en el periodo actual,

Periodo base	Periodo actual	Índices simples
X_{10}	X_{1t}	$I_1 = X_{1t}/X_{10}$
X_{20}	X_{2t}	$I_1 = X_{2t}/X_{20}$
·	·	·
·	·	·
·	·	·
X_{10}	X_{it}	$I_1 = X_{it}/X_{10}$
·	·	·
·	·	·
·	·	·
X_{no}	X_{nt}	$I_n = X_{nt}/X_{no}$

Se definen los siguientes índices complejos no ponderados.

Media Geométrica de índices simples

$$I_{t/0}^G = (I_1 \times I_2 \times \dots \times I_N)^{1/N}$$

Media armónica de índices simples,

$$I_{t/0}^H = N / \sum_{i=1}^N 1/I_i$$

Media agregativa de índices simples.

$$I_{t/0}^A = x_{1t} + x_{2t} + \dots + x_{Nt} / x_{10} + x_{20} + \dots + x_{N0} = \sum_{i=1}^N x_{it} / \sum_{i=1}^N x_{i0}$$

Ejemplo

Artículos	Precio		Índices simples	
	1970	1971	1970	1971
Leche	10	12	100	120
Queso	15	20	100	133.3
Mantequilla	80	80	100	100

Media aritmética

$$I_{1971/1970} = 120 + 133.3 + 100 / 3 = 117.76$$

Media geométrica

$$I_{1971/1970}^G = (120 \times 133.3 \times 100)^{1/3} = 116.90$$

Media armónica

$$I_{1971/1970}^H = 3 / (1/120 + 1/133.3 + 1/100) = 116.12$$

Media agregativa

$$I_{1971/1970}^A = (12+20+80) / (10+15+80) * 100 = 106.67$$

Resumiendo

Cuando se realiza una comparación entre los valores de una sola magnitud se obtienen índices simples, En cambio, si se trabaja con más de una magnitud a la vez, se habla de índices complejos. En los dos casos se comparan siempre dos situaciones, una de las cuales se considera como referencia. Cuando se trata de comparaciones temporales, a la situación inicial, se le conoce como periodo base o referencia, mientras que el periodo objeto de comparación se denomina corriente o actual. Para elaborar un número índice de carácter simple, se asigna al periodo que es objeto de referencia el valor 100, de esta manera los números índices de las distintas observaciones posteriores, no son otra cosa que porcen-

tajes de cada valor con respecto al de la referencia. Dentro de los índices complejos se distingue entre índices ponderados y no ponderados, según el peso que se le dé a los distintos valores⁹.

Índices aplicados al comercio exterior

Existe una gran diversidad de indicadores del comercio exterior, por lo que es preciso establecer una tipología que clarifique las funciones de cada uno. Un primer grupo, que denominamos indicadores agregados, analizan de forma global la posición comercial de un país. Dentro de este grupo se encuadran las medidas de apertura comercial, que reflejan el grado de inserción de una economía en los mercados internacionales. Forman parte también de este grupo los indicadores de cobertura, que informan sobre los desequilibrios en la balanza de pagos. Adicionalmente, se incluyen los indicadores de la concentración (por productos) del comercio exterior y, por último, los indicadores de la distribución geográfica del comercio.

Un segundo bloque de indicadores del comercio internacional serían los que denominamos indicadores sectoriales, que analizan la posición comercial por productos. En una primera aproximación existen dos grandes grupos: los indicadores de comercio intraindustrial y los indicadores de comercio interindustrial. Estos últimos pueden dividirse, a su vez, en indicadores de ventaja comparativa revelada o especialización comercial e indicadores de competitividad¹⁰.

Los indicadores agregados del comercio¹⁰

Los indicadores agregados del comercio exterior de un país han estado sujetos a una controversia menos intensa en cuanto a su formulación e interpretación que los indicadores sectoriales. Esto es así debido a que los primeros, que tienen un carácter eminentemente descriptivo, carecen generalmente de las connotaciones teóricas de los segundos.

La apertura exterior Los indicadores de la apertura exterior miden el grado en que una economía está integrada con el resto del mundo. El índice de apertura más común relaciona el conjunto del comercio exterior de un país con el tamaño

de su economía. Los datos del comercio exterior pueden incluir solamente a las mercancías, o bien tener en cuenta el conjunto de los bienes y servicios.

$$YM+X=I \text{ como así, } YM+X=I$$

Donde X_b y M_b , son las exportaciones e importaciones de bienes, respectivamente; X_s y M_s son las exportaciones e importaciones de bienes y servicios; mientras que Y es el PNB o PIB². Un problema asociado con este tipo de indicadores es la sobreestimación de la tasa de apertura cuando existen importantes reexportaciones de productos. En dicho caso sería relevante poder medir el contenido de valor añadido local de las exportaciones, o el contenido en importaciones de la demanda interna

El indicador de apertura comercial mencionado se enfrenta a problemas de interpretación y de delimitación de los umbrales a partir de los cuales considerar que una economía se encuentra relativamente abierta o cerrada. La apertura comercial está relacionada con diversas características estructurales (tamaño poblacional, PIB, nivel de desarrollo, estructura sectorial, políticas comerciales, etc.), por lo que el análisis de las causas de las diferencias internacionales de apertura es una tarea compleja. Por ello, Syrquin y Chenery (1989) desarrollaron un índice en el que eliminan el efecto del tamaño poblacional y el PNB por habitante. El ajuste que realizan consiste en restarle a las exportaciones reales las exportaciones esperadas, entendiendo éstas como las exportaciones que le corresponderían teóricamente al país (obtenidas a través de modelos de regresión) en función de sus características estructurales.

La tasa de cobertura

La tasa de cobertura mide el grado en que las exportaciones cubren a las importaciones. Esta tasa puede calcularse tanto para el caso del comercio de bienes, como del comercio de bienes y servicios, e incluso para el conjunto de las transacciones por cuenta corriente. En este último caso, se mide el grado en que los pagos por cuenta corriente (importaciones de bienes y servicios, las rentas y las transferencias) son cubiertos por los ingresos por cuenta corriente (exportaciones de bienes y servicios, rentas y transferencias)

$$X_b/M_b = X_{bs}/M_{bs} = I_{cc}/P_{cc}$$

Donde X_b y M_b son, respectivamente, las exportaciones e importaciones totales de mercancías; X_{bs} y M_{bs} son las exportaciones e importaciones de bienes y servicios; mientras que I_{cc} son los ingresos por cuenta corriente y P_{cc} son los pagos por dicho concepto. La necesidad de utilizar estos tres indicadores se debe a que los desequilibrios en el comercio de bienes frecuentemente se compensan con los servicios, las rentas y/o las transferencias.

Índice de precios al consumidor

Todos los días unos precios suben, otros permanecen constantes y otros bajan. Como hay millones de bienes y servicios y por lo tanto de precios, no podemos analizar la economía teniendo en cuenta todos estos detalles, sino que tenemos que utilizar una medida del nivel medio de precios¹¹.

El índice de precios al consumidor (IPC) mide la evolución del costo promedio de una canasta de bienes y servicios representativa del consumo final de los hogares, expresado en relación con un período base. La variación porcentual del IPC entre dos periodos de tiempo representa la inflación observada en dicho lapso. Las variaciones del nivel precios revelan si los precios están subiendo o bajando en promedio. El nivel de precios se mide generalmente a través de un índice de precios que se puede definir como el cociente entre el costo monetario de un conjunto dado de bienes y servicios (canasta básica o representativa) en un periodo dado y su costo en un determinado período base multiplicado por 100. Al involucrar los precios de los bienes y servicios de una canasta básica de los consumidores, hacemos referencia a un Índice de Precios al Consumidor (IPC).

Para medir el grado en que los precios varían, también existen otros índices de precios, por ejemplo, el índice de precios al por mayor o al productor. A través de éste se miden las variaciones de los precios ofrecidos al productor en todas las etapas del proceso productivo, de ese modo se tienen las categorías de bienes finales, materias primas y materias intermedias, cada una de las cuales también se divide en subcategorías. Este es considerado un índice importante,

sobre todo para observar la tendencia futura de los precios al consumidor, ya que detecta los aumentos de los precios desde el inicio del proceso productivo.

Definición de inflación Cito (...) De Pablo (1979, 147) la inflación es el aumento sostenido en el nivel general de los precios”

De esta definición surge que la inflación es un fenómeno que reúne tres características.

- Aumento de precios y no precios altos.
- Aumento de precios persistente o recurrente, lo que se opone a cambios únicos o aislados de precios.
- Generalidad del aumento de los precios, prácticamente todos los precios tienen que incrementarse.

Se define también por otros autores como un proceso económico caracterizado por alzas generalizadas y sostenidas de precios en el tiempo. Por alzas generalizadas de precios se entiende que aumentan todos los precios. Así, los precios de los bienes y servicios, el precio del servicio del trabajo. En otras palabras, sueldos y salarios. En adición, sube también el precio de las monedas extranjeras, es decir, el bolívar se deprecia frente al dólar, el marco alemán y el yen japonés, lo que significa que debemos entregar mayor número de bolívares para comprar las monedas de otros países. El otro elemento clave de la definición es el carácter de alzas de precios sostenidas en el tiempo. En otras palabras, hay que distinguir entre inflación y alzas puntuales de precios. Por ejemplo, si se decreta un aumento de sueldos y salarios, se encarecen los costos de producción lo cual se traducirá en mayores precios. Sin embargo, los precios, una vez ajustados a estos mayores costos, no tienen por qué seguir subiendo. Por el contrario, cuando hay inflación los precios suben indefinidamente sin aparentemente poder abatir la escalada, dando la impresión de encontrarnos inmersos en un proceso que se autoalimenta indefinidamente¹⁴.

La inflación ocurre cuando todos o casi todos los precios suben, como empujados por una fuerza que actuase sobre su nivel general, y cuando estos aumentos se hacen recurrentes, como si nunca fueran a detenerse. No hay infla-

ción cuando un grupo de bienes o servicios aumenta de precio con relación a los otros, como en el caso de una mala cosecha que, provocando escasez, hace subir el precio de algunos rubros alimenticios; no acostumbramos a hablar de inflación, tampoco, cuando nos referimos a un aumento generalizado pero que no se sostiene en el tiempo, como en el caso de lo que ocurría en Venezuela hasta mediados de los años setenta. La inflación es algo más general, es una especie de enfermedad de la economía a la que nadie puede escapar porque obliga de hecho a todos los actores económicos, a quienes venden o compran, a los trabajadores y a los empresarios, a subir los precios de lo que venden para no quedarse detrás. Quien no lo hace corre el riesgo de sufrir fuertes pérdidas, de ver reducidos sus ingresos o de hacer quebrar su negocio¹⁴.

Los datos referentes a la inflación nos proporcionan una valiosa información por diversos motivos:

- La inflación se utiliza como referencia para explicar numerosas variables económicas algunas tan importantes como el crecimiento del salario del trabajador, ya que los sindicatos negocian este incremento con el crecimiento de la inflación.
- Sirve a los gobiernos para valorar su política monetaria y comprobar si la oferta monetaria puesta en circulación es suficiente.
- La inflación excesiva es peligrosa, ya que si no se puede prever lo que pasará en el futuro, los agentes económicos no podrán tomar las decisiones más adecuadas.
- La inflación y el Índice de Precios al Consumidor

El Índice de Precios al Consumidor (IPC) no es lo mismo que la tasa de inflación, o simplemente la inflación, como se le llama corrientemente. El IPC es un promedio ponderado de los precios de ciertos bienes y servicios que determinadas instituciones en nuestro caso, el BCV calculan regularmente, mediante encuestas a comercios y empresas. La tasa de inflación es el aumento que el IPC presenta entre dos períodos concretos expresada en términos de porcentaje. Ejemplificamos; en diciembre de 1985 el índice era, digamos, de 361 Bs., y en diciembre de 1986 resultaba de 403 Bs. la diferencia ($403 - 361 = 42$), expresada como tanto por ciento (o sea, $42/36 \times 100$) es la inflación que ha habido durante

1986, en este caso 11,6%.

Se plantea que cuando la inflación baja ello no significa, por lo tanto, que el IPC haya disminuido. Ejemplificamos, la tasa de inflación baja de 3% a 2% mensual eso quiere decir que lo que se ha reducido es la velocidad en el aumento del IPC: este ha aumentado entre un marzo y abril en un 2%, algo menos que lo que había hecho entre febrero y marzo (3%), pero, sin embargo, sigue creciendo.

Cálculo e interpretación de la inflación. La tasa de inflación es la tasa a la que crece el nivel general de precios (generalmente el IPC). Esta tasa de crecimiento podría medirse respecto a cualquier momento anterior que se quiera:

$$\pi_t = \frac{P_t - P_{t-k}}{P_{t-k}} \cdot 100$$

Índices de cotizaciones de bolsa de valores

Desde hace tiempo los economistas han estudiado e intentado comprender los movimientos de los precios en la bolsa de valores, debido a que las inversiones en bolsa están sujetas a riesgos, los rendimientos son variables y su existencia es incierta. La predicción de la bolsa de valores es un tema de interés, en particular para quienes invierten en ella. Sería muy provechoso poder predecir la tendencia y, si fuera posible, el precio de las acciones, ya que con esta información los inversionistas podrían realizar movimientos apropiados y así ganar dinero.

Por lo tanto, predecir un índice de la bolsa de valores representa un gran reto y en las últimas dos décadas ello ha sido objeto de muchos estudios dadas las aplicaciones comerciales que tiene. Numerosos métodos han sido propuestos para brindar predicciones más precisas a los inversores. Algunos de estos estudios han utilizado modelos autorregresivos, promedios móviles, Arima (Reddy, 2010), regresión múltiple (Chang, Yeung y Yip, 2000), algoritmo genético (Kim y Han, 2000), redes neuronales artificiales (Chena, Leung y Daouk, 2003), suavizado exponencial, métodos lineales y no lineales (Zemke, 1998), entre otras.

En este artículo se hace una revisión bibliográfica acerca de los principales métodos que se han utilizado para predecir índices bursátiles¹⁶.

Las bolsas de valores son consideradas como uno de los barómetros más sensibles de la situación económica del país y los índices bursátiles en particular son seguidos atentamente por los inversores.

Los índices bursátiles representan una ponderación de un conjunto de valores o acciones que cotizan en un mismo mercado y reflejan en resumidas cuentas las variaciones promedio de los elementos que componen al índice. En términos generales cuando los principales índices bursátiles registran ganancias, en el mercado reina optimismo, y por lo general los inversores tienden a comprar activos de mayores riesgos y monedas como euros y libras esterlinas.

Sin embargo, cuando los principales índices mundiales registran pérdidas, los inversores tienden a refugiarse en activos seguros y en monedas como el dólar estadounidense, el yen japonés y el franco suizo, ya que son activos que mantienen un status de refugio seguro y de resguardo sobre todo cuando la aversión al riesgo predomina en la bolsa.

Cada bolsa de valores tiene sus propios índices bursátiles y por lo general se diferencian en cinco aspectos básicos como ser la muestra de activos que toma el índice (cantidad de empresas que lo componen), el tipo de ponderación que se da a cada activo dentro del índice (factores como la capitalización bursátil y el volumen de contratación), la fórmula matemática utilizada para calcular el índice (si se utilizan medias geométricas o medias aritméticas, etc), la fecha de referencia del índice (año de toma del valor base 100), y en la forma de realizar ajustes al índice (existen índices que se ajustan por pago de dividendos o ampliaciones de capital, etc)

Causas de la inflación

Hay distintas explicaciones sobre causas de inflación. Hay diversos tipos de procesos económicos que producen inflación, cada explicación trata de dar cuenta de un proceso generador de inflación diferente.

Existen al menos tres tipos de inflación:

- Inflación de demanda (demand pull inflation), cuando la demanda general de bienes se incrementa, sin que el sector productivo haya tenido tiempo de adaptar la cantidad de bienes producidos a la demanda existente
- Inflación de costos (cost push inflation), cuando el costo de la mano de obra o las materias primas se encarece, y en un intento de mantener la tasa de beneficio los productores incrementan los precios
- Inflación auto-construida (build-in inflation), ligada al hecho de que los agentes prevén aumentos futuros de precios y ajustan su conducta actual a esa previsión futura. Es la inflación generada por las expectativas de inflación futura, círculo vicioso
- El esquema más aceptado sobre la causa de la inflación es la que indica simplemente que la inflación la promueve la expansión de la masa monetaria a una tasa superior a la expansión de la economía.
- Siguiendo esta teoría: al incrementarse la masa monetaria, la demanda por bienes aumenta y -si esta no viene acompañada en un incremento en la oferta de bienes- la inflación surge.
- Siguiendo esta línea de pensamiento, el control de la inflación descansa en la prudencia fiscal y monetaria; es decir el gobierno debe asegurarse que no sea muy fácil obtener préstamos, ni tampoco debe endeudarse él mismo significativamente.
- Por lo tanto, este enfoque monetarista resalta la importancia de controlar los déficits fiscales y las tasas de interés, así como la productividad de la economía.

Entre los principales índices de las bolsas más importantes del mundo tenemos:

Dow Jones Industrial Average - El Promedio Industrial de Dow Jones es un promedio de precios ponderado de las 30 empresas más importantes que tienen base en Estados Unidos. Incluye a empresas como Alcoa, Coca Cola, Microsoft, Boeing, Amex, Citigroup, General Motors y Pfizer. Nasdaq Composite - El Índice Compuesto NASDAQ mide todos los valores domésticos e internacionales que cotizan en el Mercado de Valores del Nasdaq. El índice incluye más de

3.000 empresas, más que en la mayoría de los índices bursátiles. Este índice es seguido muy de cerca en Estados Unidos como evolución del sector tecnológico y crecimiento, pero no es un índice exclusivo de Estados Unidos ya que incluye a empresas de todo el mundo.

S&P 500 - Es un índice bursátil que contiene los valores de 500 empresas de gran capitalización, todas ellas de los Estados Unidos. Se considera un barómetro de la economía de los Estados Unidos.

IBEX 35 - Es el índice bursátil de referencia de la Bolsa de Madrid, la principal bolsa de valores de España. Es un índice ponderado de los 35 valores españoles con mayor liquidez en el mercado de Madrid. Está compuesto por compañías como Iberia, BBVA, Repsol YPF, FCC, y Telefónica.

DAX (DeutscherAktienIndex) - Es un índice bursátil que representa a 30 de las empresas alemanas más grandes y con mayor liquidez que cotizan en la Bolsa de Frankfurt. Se incluyen empresas como Adidas, Basf, Bayer, Deutsche Bank, Merck, SAP, Siemens y los fabricantes de automóviles Volkswagen y BMW.

CAC 40 (ContinuousAssistedQuotation) - Es el índice bursátil de Francia que sigue el curso de los 40 mayores valores de Francia que cotizan en el mercado bursátil de la Bolsa de París ahora Euronext Paris. Se incluyen empresas como Renault, Air France, Michelin, Alcatel-Lucent, L'Oreal y Vivendi.

FTSE 100 - Es el índice bursátil que representa a 100 de las empresas británicas de mayor tamaño que cotizan en la Bolsa de Londres. Incluye a empresas como British Airways, British Petroleum, HSBC, Lloyds TSB, Tesco y Unilever.

Nikkei 225 - Es el índice bursátil más popular en Japón que está compuesto por los 225 activos más líquidos que cotizan en el mercado bursátil de la Bolsa de Tokio. Se incluyen empresas como Honda, Mazda, Toyota, Hitachi, Panasonic y Sony¹⁸.

Deflactar

Transformar una magnitud económica expresada en términos monetarios a precios corrientes, en otra magnitud expresada también en términos monetarios a precios del año base. Este procedimiento permite eliminar del valor de dicha magnitud el efecto de la inflación o subida de precios.

La deflación de series monetarias consiste en eliminar el efecto que los cambios en los precios de los bienes tienen sobre las series de valores.

Cuando queremos conocer la evolución de una serie de valores a lo largo del tiempo, por ejemplo, beneficios de una empresa, producción de una industria, salarios de los empleados de una empresa, ingresos de los hogares, etc., nos encontramos habitualmente con los valores están en unidades monetarias de cada periodo, esto es, los valores se refieren a unidades monetarias corrientes. Esto va a hacer que los valores no sean directamente comparables puesto que las alteraciones de los precios de un periodo a otro confieren distinto poder adquisitivo a las unidades monetarias. En otras palabras, el efecto de la inflación (o deflación) modifica la capacidad de compra del dinero.

Para conocer los cambios reales experimentados por la serie a lo largo del periodo de interés tendremos que expresar todos los valores de dicha serie en unidades monetarias de un mismo periodo, es decir, en unidades monetarias constantes.

Los valores expresados en unidades monetarias corrientes se conocen como valores nominales.

Los valores expresados en unidades monetarias constantes se conocen como valores reales.

El paso de una serie de unidades monetarias corrientes a unidades monetarias constantes se denomina deflación. Para deflacionar una serie basta con dividir la serie original por un índice de precios adecuado que elimine la influencia de los precios. Este índice recibe el nombre de deflacionador. (Aunque la de-

nominación correcta en castellano es deflacionador, y la acción, deflacionar, es habitual en los medios de comunicación, e incluso en los medios académicos, utilizar el anglicismo deflactor y deflactar, respectivamente).

La indexación es una técnica para ajustar pagos de ingresos mediante un índice de precios, para mantener el poder adquisitivo del público luego de la inflación, mientras desindexación es la anulación de la indexación.

Se conceptualiza la indexación como el acomodo de los precios al consumidor ya que se tiene que ir emparejando conforme aumente el salario mínimo, siempre deben andar a la par no puede uno estar más bajo del otro, debe existir un equilibrio. Cambio de base y empalme, diciendo que este ejercicio se trabaja de la siguiente manera, como nos dice es cambio de base, en vez de tomar la primera base, esta vez se tomara la última base ó sea la del año anterior para tener estadísticos con mayor exactitud y más recientes, esta se llama base variable porque va ir cambiando conforme transcurra el tiempo.

Desde un punto de vista macroeconómico, existen cuatro categorías principales de indexación: indexación salarial, indexación de tasas de instrumentos financieros, indexación de tasas impositivas e indexación de tipos de cambio. Los primeros tres están indexados por inflación. El último generalmente está indexado a una moneda extranjera, principalmente el dólar estadounidense. Cualquiera de estos tipos diferentes de indexación se puede revertir (desindexación).

La indexación de la moneda o tasa de cambio a menudo se refiere a un país que vincula su moneda al dólar estadounidense. En otras palabras, el banco central de ese país compraría o vendería dólares para mantener un tipo de cambio estable con el dólar. Tal política ha sido adoptada por varios países asiáticos, incluida China. De no ser por la vinculación mencionada, las monedas de estos países aumentarían frente al dólar como resultado del déficit crónico de la cuenta corriente de los Estados Unidos con dichos países. Sin embargo, los países asiáticos tienen un interés económico determinado en mantener alta la demanda estadounidense de sus exportaciones.

Ahí es donde entra la vinculación de su moneda con el dólar estadounidense.

A menudo, la vinculación realizada por los bancos centrales es bastante discreta y no se revela en ninguna declaración formal de política. La vinculación también puede ser bastante elástica. Un banco central mantendrá un tipo de cambio dentro de un rango considerado aceptable en lugar de a un nivel específico. Con el tiempo, el rango aceptable puede ampliarse o reducirse dependiendo de la dependencia económica general de ese país de las exportaciones para impulsar el crecimiento. Por lo tanto, es difícil observar claramente la desindexación de una moneda.

Objetivos del análisis de regresión y correlación

La correlación y regresión son ideas básicas en estadística, utilizadas en las ciencias y un gran número de campos de la actividad humana. Estas técnicas amplían la idea de dependencia funcional a situaciones aleatorias.

Se define como correlación, a una medida en la cual dos o más variables encuentran relaciones de interdependencia entre sí.

El estudio correlacional, es aquel realizado en la investigación científica donde existe manipulación específica de las variables de estudio, a través de un procedimiento de selección. Estas variables ya definidas en este tipo de estudios difieren de las utilizadas en los estudios experimentales, donde las variables son “creadas” por el investigador, manipulándose en forma directa.

Las relaciones existentes entre las variables de estudio pueden ser expresadas mediante una relación directa o mediante el hallazgo de una relación inversa o negativa entre las variables. Por lo tanto, la correlación es la co-variación entre las variables, por lo que el término correlación y covarianza significan lo mismo, en el entendido, del análisis de la variación de las variables de estudio.

Si bien los estudios correlacionales, son de tipo observacional al igual que los estudios descriptivos, estos últimos miden con precisión las variables individuales y los conceptos de las mismas, y no la relación existente entre ellas, pueden ser realizados en ambientes naturales o laboratoriales siempre y cuando la manipulación de las variables no sea directa.

El análisis de regresión permite establecer la relación funcional o ecuación matemática que relaciona las variables, así como la fuerza de esa relación.

En la terminología de la regresión, la variable que se va a predecir se llama dependiente, a explicar, o endógena. La o las variables que se usan para predecir el valor de la variable dependiente se llaman independientes, explicativas o exógenas.

En general, existen cuatro posibles formas en que las variables se pueden relacionar, a saber: Relación lineal directa, relación lineal inversa, relación no lineal directa y relación no lineal inversa, cuya estructura formal y funcional, permite dilucidar con objetividad las actividades orientadas a decidir qué ecuación se debe emplear, cuál ha de ser la ecuación que mejor se ajusta a los datos y cómo debe validarse la significancia estadística de los pronósticos realizados.

Regresión lineal: Curva de ajuste

Regresión lineal. Permite determinar el grado de dependencia de las series de valores X e Y, prediciendo el valor y estimado que se obtendría para un valor x que no esté en la distribución. Existen diferentes tipos de regresión lineal que se clasifican de acuerdo a sus parámetros: Regresión lineal simple La regresión lineal simple se basa en estudiar los cambios en una variable, no aleatoria, afectan a una variable aleatoria, en el caso de existir una relación funcional entre ambas variables que puede ser establecida por una expresión lineal, es decir, su representación gráfica es una línea recta. Es decir, se esta en presencia de una regresión lineal simple cuando una variable independiente ejerce influencia sobre otra variable dependiente.

Ejemplo: $Y = f(x)$

Regresión lineal múltiple

La regresión lineal permite trabajar con una variable a nivel de intervalo o razón, así también se puede comprender la relación de dos o más variables y permitirá relacionar mediante ecuaciones, una variable en relación a otras varia-

bles llamándose Regresión múltiple. O sea, la regresión lineal múltiple es cuando dos o más variables independientes influyen sobre una variable dependiente. Ejemplo : $Y = F(x, w, z)$

Aplicaciones de la regresión lineal

Líneas de tendencia

Una línea de tendencia representa una tendencia en una serie de datos obtenidos a través de un largo período. Este tipo de líneas puede decir si un conjunto de datos en particular (como por ejemplo, el PBI, el precio del petróleo o el valor de las acciones) han aumentado o decrementado en un determinado período. Las líneas de tendencia son generalmente líneas rectas, aunque algunas variaciones utilizan polinomios de mayor grado dependiendo de la curvatura deseada en la línea.

Métodos de mínimos Cuadrados. Para estudiar el comportamiento de una serie de datos obtenidos empíricamente, constituidos por puntos dados mediante pares ordenados de números, asociados con los valores de dos variables, es necesario contar con una función, que exprese analíticamente la relación funcional que guardan las variables en cuestión.

La interpolación: se caracteriza por suponer que los datos que intervienen en el problema son exactos; por lo cual en la construcción de la función de interpolación se exige que la misma satisfaga todos y cada uno de los valores que constituyen los datos.

El ajuste se supone que los datos ingresados están afectados en cierto grado de errores debido al modelado, por lo que, no resulta indispensable que la CURVA DE AJUSTE correspondiente, pase exactamente por los puntos que representan los datos, sino que, en promedio la aproximación sea óptima de acuerdo a un cierto y determinado criterio, denominado CRITERIO DE AJUSTE. El iniciador de estos procedimientos fue Gauss, quien desarrollo el tan conocido método de los mínimos cuadrados.

Método de los mínimos cuadrados

Es una técnica de Análisis Numérico en la que, dados un conjunto de pares (o ternas, etc), se intenta encontrar la función que mejor se aproxime a los datos (un “mejor ajuste”). En su forma más simple, intenta minimizar la suma de cuadrados de las diferencias ordenadas (llamadas residuos) entre los puntos generados por la función y los correspondientes en los datos.

Fortaleza

- Es objetivo, sólo depende de los resultados experimentales.
- Es reproducible, proporciona la misma ecuación, no importa quién realice el análisis.
- Proporciona una estimación probabilística de la ecuación que representa a unos datos experimentales
- Proporciona intervalos pequeños de error.

Restricciones

- Sólo sirve para ajustar modelos lineales
- Requiere tener, al menos, diez mediciones bajo las mismas circunstancias experimentales.
- Tales resultados deben estar descritos por una distribución de probabilidad conocida. La más común es la distribución normal o gaussiana
- Se requiere de algún equipo de cálculo, de lo contrario, es muy engorroso.

Regresión no lineal

Si la relación no es lineal, pueden transformarse los valores de una o ambas variables para intentar linealizarla. Si no es posible convertir la relación en lineal, puede comprobarse el grado de ajuste de una función polinomial más compleja. La función polinomial más sencilla es la cuadrática $y = w_0 + w_1 x_1 + w_2 x_2$ que describe una parábola, pero puede usarse una función cúbica u otra de un orden aun mayor (orden k) capaz de conseguir un ajuste casi perfecto a

los datos.

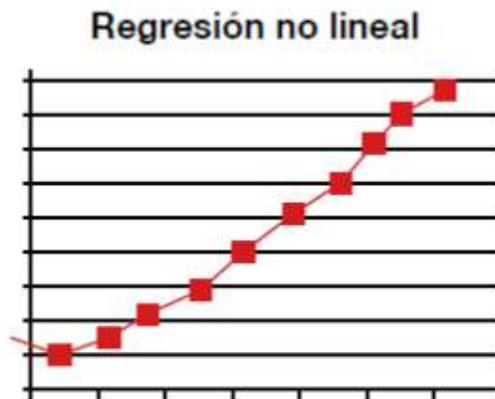
$$Y_i = X_i w = w_0 + w_1 x_1 + w_2 x_{2i} + \dots + w_k x_i^k$$

$$\text{Para } X_i = (1, x_i, x_{2i}, \dots, x_{ki})$$

Las fronteras de decisión no lineales permiten representar conceptos más complejos al ajustarse más a los datos, no obstante este sobreajuste implica también inconvenientes, instancias de entrenamiento ruidosas (outliers) son también sobre ajustadas, desplazando estas fronteras hacia esas instancias equivocadas y ocasionando así confundir al sistema de predicción a la hora de predecir nuevas entradas. Este sobreajuste (overfitting) es un problema muy común y produce un modelo que no es capaz de generalizar. Normalmente, fronteras de decisión muy complejas producen sobreajuste, no funcionando adecuadamente con nuevas instancias.

Los modelos no lineales se originan cuando un investigador obtiene, por el desarrollo de una teoría o por otra situación, una relación funcional en la que los parámetros son no lineales.

La regresión lineal suele conseguir fronteras de decisión más correctas y menos artificiales que la regresión no lineal. A pesar de producir mayores errores con los Ejemplos de entrenamiento, tiene mayor capacidad de generalización y se comporta mejor ante nuevos ejemplos a predecir.



Una regresión exponencial es el proceso de encontrar la ecuación de la fun-

ción exponencial que se ajuste mejor a un conjunto de datos.

La ecuación de la fórmula es: $y = a \cdot b^x$

Una línea de tendencia exponencial es una línea curva que es muy útil cuando los valores de los datos aumentan o disminuyen a intervalos cada vez mayor.

No es posible crear una línea de tendencia exponencial si los datos contienen valores cero o negativos.

Regresión mínimo cuadrática no-lineal

La regresión mínimo-cuadrática puede plantearse de forma que la función de ajuste se busca no sea una función lineal. El planteamiento general sería similar, aunque obviamente habría que minimizar el cuadrado de los residuos entre los datos originales y los valores teóricos obtenibles a través de la función no-lineal considerada.

Serie cronológica

Una serie cronológica, está dado por un conjunto de observaciones que están ordenadas en el tiempo, y que estas pueden representar el cambio de una variable ya sea de tipo económica, física, química, biológica... a lo largo su historia.

El objetivo del análisis de una serie cronológica es el conocimiento de su patrón de comportamiento, para así poder prever su evolución en el futuro cercano, suponiendo por supuesto que las condiciones no variaran significativamente.

El análisis de series cronológicas comprende métodos que ayudan a interpretar este tipo de datos extrayendo información representativa, tanto referente a los orígenes o relaciones subyacentes como a la posibilidad de extrapolar y predecir su comportamiento futuro.

Componentes de las series cronológicas.

Las series cronológicas se suelen presentar por medio de una ecuación matemática que describa los valores de la variable observada como una función del tiempo, es decir ($Y = f(t)$). Al representar gráficamente la información en un sistema de coordenadas en el eje de las ordenadas se ubica la variable y en el de las abscisas el tiempo. Esta representación gráfica es difícil para detectar los movimientos de la serie, los cuales son causados por una variedad de factores que pueden ser económicos, naturales, institucionales o culturales.

Existen diferentes métodos para analizar una serie de tiempo, siendo uno de ellos el modelo de descomposición, el cual considera que la serie está compuesta de cuatro patrones básicos:

- La tendencia (T)
- Variaciones estacionales (S)
- Variaciones cíclicas (C)
- Variaciones irregulares o aleatoria (I)

Indica la marcha general y persistente del fenómeno observado, es una componente de la serie que refleja la evolución a largo plazo. Por ejemplo, la tendencia creciente del índice de reciclado de basuras en los países desarrollados, o el uso creciente de Internet en la sociedad, independientemente de que en un mes concreto en un país, por determinadas causas, haya una baja en la utilización de Internet.

Lo que mide la tendencia es la variación promedio de la variable por unidad de tiempo. Esta tendencia se describe mediante una recta o algún tipo de curva lisa.

Variaciones estacionales (S)

Es el movimiento periódico de corto plazo. Se trata de una componente causal debida a la influencia de ciertos fenómenos que se repiten de manera periódica en un año (las estaciones), una semana (los fines de semana) o un día (las horas punta) o cualquier otro periodo. Recoge las oscilaciones que se producen en esos periodos de repetición.

Variación cíclica

Es el componente de la serie que recoge las oscilaciones periódicas de amplitud superior a un año. Movimientos normalmente irregulares alrededor de la tendencia, en las que, a diferencia de las variaciones estacionales, tiene un período y amplitud variables, pudiendo clasificarse como cíclicos, cuasi cíclicos o recurrentes.

Variaciones irregulares o aleatoria (I)

Se deben a razones aleatorias esporádicas y por tanto impredecibles. No obstante estos sucesos se pueden reconocer e identificando fácilmente. Las variaciones aleatorias son de dos clases:

Variaciones provocadas por acontecimientos especiales , como elecciones , guerras , inundaciones etc Variaciones aleatorias o por casualidad, cuyas causas no se pueden señalar en forma exacta. Promedios móviles.

Las medias móviles son indicadores de tendencias que se utilizan para analizar los datos anteriores con el fin de crear una serie de medias de diferentes subconjuntos de datos de precios. En otras palabras, las medias móviles examinan los precios históricos y forman medias de esos precios, suavizándolas durante un período de tiempo; se clasifican como indicadores rezagados, debido al hecho de que utilizan datos anteriores para trazar la nueva línea de precios. Tienden a dar señales después de que la nueva tendencia o el movimiento importante de los precios ha comenzado.

Por sí solos, no son predictores muy precisos del movimiento de los precios. Deben combinarse (es decir, una media móvil de corto plazo y una media móvil de largo plazo) para producir una señal, o combinarse con otros indicadores que puedan medir el impulso de la acción de los precios, como los osciladores.

Las medias móviles son capaces de utilizar el precio abierto, el precio de cierre, el precio alto y el precio bajo como precios de referencia para alisar los

datos de precios. Hay varios tipos de promedios móviles: Promedio móvil simple (SMA), Promedio Móvil Ponderado (WMA).

Los promedios móviles su utilidad en el comercio son los de corto plazo (como 50 SMA) y un promedio móvil de largo plazo (como 200SMA), para producir una señal cruzada. El promedio móvil de corto plazo responderá más rápido que su contraparte de largo plazo al reaccionar a los cambios de precios. La señal aquí es esperar a que el cruce del MA cortó sobre el MA largo en cualquier dirección para apuntar a la nueva tendencia. Algunos también añadirán otros indicadores en un intento de filtrar las señales de retardo y captar los movimientos tempranamente.

BIBLIOGRAFÍA

NUMERO ÍNDICES Y ANÁLISIS

DE REGRESIÓN SIMPLE



www.mawil.us

1. BayarreVea H , Oliva Perez M , HorsfordSaing M , Ranero Aparicio V , Coutin Marie G , Diaz Llanes G . Tema I .Estadística Descriptiva. Metodología de la Investigación en APS. 2004.
2. Tema 4. Números índice. Universidad autónoma del estado de México. Facultad de Contaduría y administración [citado 2 Mar 2019]. Disponible en :http://www.seduca2.uaemex.mx/ckfinder/uploads/files/u3tema_4_numeros__nd.pdf
3. Carmona F . Numero de Índices. 2001[citado 2 Mar 2019]. Disponible en :<http://www.ub.edu/stat/docencia/Mates/indices.pdf>
4. López Cachero M . Fundamentos y métodos estadísticos .Editorial PirámideSBN: 9788436804256 Numero de índices simples:
5. Fernández S. Estadística Descriptiva: Números Índices. 2013 [citado 2 Mar 2019]. Disponible en :<http://www.fuenterrebollo.com/Economicas2013/indices-teoria.pdf>
6. Unidad 3. Construcción de números índice y aplicaciones al análisis económico. [citado 2 Mar 2019]. Disponible en :<http://ri.uaemex.mx/bitstream/handle/20.500.11799/32041/secme-21242.pdf?sequence=1>
7. Números índices, período base , período actual clasificación de los números índices (simples, complejos) ejemplo. [citado 2 Mar 2019]. Disponible en :<https://www.uv.es/ceaces/numindices/clasifica.htm>
8. Unidad 3. Construcción de números índice y aplicaciones al análisis económico[citado 2 Mar 2019]. Disponible en :<http://ri.uaemex.mx/bitstream/handle/20.500.11799/32041/secme-21242.pdf?sequence=1>
9. López Cachero M. Fundamentos y métodos estadísticos. Editorial Pirámide SBN: 9788436804256
10. Hernández Martín R Indicadores para el análisis del comercio exterior. Una visión panorámica. 2001.[citado 2 Mar 2019]. Disponible en :https://www.researchgate.net/publication/232076692_Indicadores_para_el_analisis_del_comercio_exterior_Una_vision_panoramica
11. El índice de precios al consumidor y la inflación[citado 2 Mar 2019]. Disponible en : 2018. [Http://nulan.mdp.edu.ar/2880/1/atucha-et-al-2018.pdf](http://nulan.mdp.edu.ar/2880/1/atucha-et-al-2018.pdf)
12. Índice de precios al consumidor (IPC). 2019 [citado 2 Mar 2019]. Disponible en :<http://www.banrep.gov.co/es/indice-precios-consumidor-ipc>
13. Índice de Precios de Consumo. 2001[citado 2 Mar 2019]. Disponible en :<http://www5.uva.es/estadmed/datos/indices/indices7.htm>

14. Hugo J. Faría y Carlos Sabino . La Inflacion1997 [citado 2 Mar 2019]. Disponible en : <http://paginas.ufm.edu/sabino/word/inflacion.pdf>
15. Gutiérrez Andrade, Osvaldo; Zurita Moreno, Andrea Sobre la inflación Perspectivas, vol. 9, núm. 3, 2006 [citado 2 Mar 2019]., pp. 81-115 Universidad Católica Boliviana San Pablo Cochabamba, Bolivia Disponible en : <https://www.redalyc.org/pdf/4259/425942413004.pdf>
16. García M ,Jalal A , Garzón L, López M Métodos para predecir índices bursátiles. 2013 [citado 2 Mar 2019]. Disponible en : <http://www.scielo.org.co/pdf/ecos/v17n37/v17n37a3.pdf>
17. Mimica E S .Los Efectos de la Inflación en la Evaluación de Proyectos de Inversión.[citado 2 Mar 2019]. Disponible en : https://www.cepal.org/ilpes/noticias/paginas/2/33602/efectos_de_inflacion.pdf
18. ¿Cuáles son las bolsas más importantes del mundo?. 2018. [citado 2 Mar 2019]. Disponible en: <https://www.ig.com/es/estrategias-de-trading/-cuales-son-las-bolsas-mas-importantesdel-mundo--181015>
19. ¿Cómo se mide la deflactación de magnitudes económicas? . 2016.[citado 2 Mar 2019]. Disponible en : http://www.diofante.com/wp-content/uploads/2016/08/COMO_SE_MIDE_Numero_1_Julio_2016.pdf
20. Deflación de series monetarias.[citado 2 Mar 2019]. Disponible en : <http://www5.uva.es/estadmed/datos/indices/indices6.htm>
21. Gea M .M , Arteaga P , Batanero, Ortiz J . Conocimiento Tecnológico sobre la Correlación y Regresión: u n estudio exploratorio con Futuros Profesores. 201 Bolema, Rio Claro (SP), 2018; v. 32, n. 60, p. 134-155
22. Bustamante C Gladys, Mendoza Quispe Carla Anahi. Estudios de Correlacion. Rev. Act. Clin. Med . [citado 2019 Feb 19]. Disponible en: http://www.revistasbolivianas.org.bo/scielo.php?script=sci_arttext&pid=S2304-37682013000600006&lng=es.
23. Estudios correlacionales. Tema 5. Introducción a la Psicología. 2013 [citado 2 Mar 2019]. Disponible en : <http://www4.ujaen.es/~eramirez/Descargas/>
24. tema5. Morales Vallejo P. Correlación y covarianza. Universidad Pontificia Comillas. Madrid. 2008:1-48
25. Sáenz Campos D.. Tinoco Mora Z. Introducción a la investigación científica. [citado 2 Mar 2019]. Disponible en : <http://www.cendeisss.sa.cr/etica/art2.pdf>.

26. Cardona Madariaga D. F , González Rodríguez J , Rivera Lozano M , Cárdenas Vallejo E H. Aplicación de la regresión lineal en un problema de pobreza. 2013. [citado 2 Mar 2019]. Disponible en :<http://www.unilibre.edu.co/revistainteraccion/volumen12/art4.pdf>
27. Regresión lineal .EcuRed .[citado 2 Mar 2019]. Disponible en :https://www.ecured.cu/Regresi%C3%B3n_lineal
28. Kelmansky D .Regresión lineal simple. 2008. [citado 2 Mar 2019]. Disponible en :http://www.dm.uba.ar/materias/analisis_de_datos/2008/1/teoricas/Teor8.pdf
29. Análisis numérico/ métodos numéricos. 2008[citado 2 Mar 2019]. Disponible en :<http://exa.unne.edu.ar/matematica/metodos/5-3-material-teorico/min-cuadrado.pdf>
30. Rivas M. Regresión no Lineal. Revista Colombiana de Estadística N^o 27 [citado 2 Mar 2019]. Disponible en :<http://www.bdigital.unal.edu.co/15397/1/10003-18133-1-PB.pdf>
31. Pereira González Análisis predictivo de datos mediante técnicas de regresión estadística. 2010 [citado 2 Mar 2019]. Disponible en :https://eprints.ucm.es/11389/1/Analisis_Predictivo_de_Datos.pdf
32. Serie cronológica. [citado 2 Mar 2019]. Disponible en: https://www.academia.edu/20808075/Series_Cronologicas
33. Series de Tiempo. Tema 3 .[citado 2 Mar 2019]. Disponible en :http://www.seduca2.uaemex.mx/ckfinder/uploads/files/u3tema_3_series_de_t.pdf
34. ¿Qué son los promedios móviles y cómo se utilizan en el comercio?. 2018. [citado 2 Mar 2019]. Disponible en :<https://www.tokens24.com/es/cryptopedia/trading/que-son-lospromedios-moviles-y-como-se-utilizan-en-el-comercio>

UNIDAD IV

TEORÍA DE LAS PROBABILIDADES Y DISTRIBUCIÓN DE PROBABILIDADES NORMAL



www.mawil.us

La probabilidad es la ciencia que trata de cuantificar los posibles resultados de un experimento en el cual está presente la incertidumbre o la aleatoriedad. La teoría de la probabilidad se usa extensamente en áreas como la estadística, la física, la matemática, la ciencia y la filosofía para sacar conclusiones sobre la probabilidad de sucesos potenciales y la mecánica subyacente de sistemas complejos¹.

Definición de eventos, y espacio muestral: Aplicaciones. Clasificación de eventos y de espacio muestral En la teoría de la probabilidad, un evento aleatorio o fuente de sucesos aleatorio es un subconjunto de un espacio muestral, es decir, un conjunto de posibles resultados que se pueden dar en un experimento aleatorio. En teoría de la probabilidad a cada evento aleatorio se le puede asignar una medida de probabilidad, y el conjunto de todos los sucesos aleatorios constituye una σ -álgebra de conjuntos¹.

Se clasifican en:

- Evento simple, siendo aquel que tiene un solo punto muestral.
- Evento compuesto, siendo aquel que tiene dos o más puntos muestrales.

Se conceptualiza espacio muestral al conjunto de todos los resultados posibles de un experimento aleatorio. El espacio muestral se denota como S .

Ejemplo: Los resultados posibles del lanzamiento de un dado.

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

Ejemplo: Los resultados posibles del lanzamiento de una moneda.

$$S = \{\text{Sello}, \text{Águila}\}$$

Los espacios muestrales se clasifican en:

- Espacio muestral discreto :son espacios muestrales cuyos elementos resultan de hacer conteos, siendo por lo general subconjuntos de los números enteros.
- Espacio muestral continuo: son espacios muestrales cuyos elementos

resultan de hacer mediciones, siendo por lo general intervalos en el conjunto de los números reales.

Ante estos conceptos es posible llegar a pensar que un evento y un punto muestral son lo mismo, pero realmente no lo son. Un ejemplo claro se puede observar en el lanzamiento del dado, un evento sería por ejemplo que salga número par, para lo cual servirían los puntos muestrales. De ahí las diferencias entre unos y otros.

Técnicas de conteo Principio aditivo y multiplicativo⁴.

Dado que la probabilidad se refiere a la potencialidad de ocurrencia de un evento, el principio aditivo se refiere a las formas que ese evento puede ser realizado. Por ejemplo, una persona que define viajar desde Santiago al Litoral Central puede hacerlo por Línea de Buses A, Línea de Buses B, Línea de Buses C, Línea de Buses D. El principio aditivo, sería que cada línea de buses representa una alternativa: $L_A = 1$; $L_B = 1$; $L_C = 1$; $L_D = 1$, (significa que cada línea de buses tiene una línea disponible al litoral central)

En el principio Aditivo sería, que la forma de llegar al punto L sería:

$$L = 1 + 1 + 1 + 1 = 4$$

La clave en el principio aditivo es buscar intrínsecamente la “0”, en el ejemplo, la persona para dirigirse al litoral central no puede utilizar todas las alternativas, tiene que utilizar una “o” la otra. Cuando se use el “o”, entonces hay que utilizar el principio aditivo.

Principio Multiplicativo

El principio multiplicativo consiste en que, si existen distintas formas de que un evento suceda, y a su vez estas distintas formas tienen subformas de realizarse, se utiliza la multiplicación: se utiliza la cantidad de formas, por la cantidad de sub formas.

Por ejemplo, en el caso anterior, dado que existen cuatro líneas de buses, suponiendo que la línea A tenga cinco buses, la línea B tenga cuatro buses, la línea C tenga dos buses y la línea D tenga ocho buses, entonces la forma de llegar a L, aplicando el principio multiplicativo, sería: $L = 1 \times 5 + 1 \times 4 + 1 \times 2 + 1 \times 8$

Entonces el resultado sería $L = 19$

Permutaciones

Para entender lo que son las permutaciones es necesario definir lo que es una combinación y lo que es una permutación para establecer su diferencia y de esta manera entender claramente cuando es posible utilizar una combinación y cuando utilizar una permutación al momento de querer cuantificar los elementos de algún evento.

La aplicación de la definición clásica de probabilidad puede presentar dificultades de aplicación cuando el espacio muestral es infinito o cuando los posibles resultados de un experimento no son equiprobables. Ej: En un proceso de fabricación de piezas puede haber algunas defectuosas y si queremos determinar la probabilidad de que una pieza sea defectuosa no podemos utilizar la definición clásica pues necesitaríamos conocer previamente el resultado del proceso de fabricación.

Para resolver estos casos, se hace una extensión de la definición de probabilidad, de manera que se pueda aplicar con menos restricciones, llegando así a la definición frecuentista de probabilidad.

Leyes de la probabilidad

Las leyes de la probabilidad se utilizan para cuantificar los patrones que se observan en fenómenos aleatorios, desde aquéllos tan simples como el lanzamiento de dados hasta los más complejos, como la trayectoria que sigue una partícula en un acelerador de partículas, la evolución de una epidemia, el crecimiento de una población, el comportamiento de mercados financieros, el crecimiento de galaxias, etc. Usando las leyes de la probabilidad se pueden evaluar,

mediante cuantificación, las consecuencias esperadas con los diversos resultados posibles en la realización de un fenómeno aleatorio. Esto se lleva a cabo asignando un valor numérico a cada posible resultado, es decir, su utilidad, multiplicándolo por el porcentaje de veces en que ocurre, el cual lo da la ley de probabilidad asociada al fenómeno, y sumar todos los valores obtenidos. Dicho de otro modo, calculando la esperanza matemática.

Por ejemplo, un juego de apuestas en el que se lanza un dado equilibrado y si el resultado es 3 nuestra ganancia será de 1 peso, mientras que de resultar otro número nuestra pérdida será de 1 peso, por lo que la utilidad esperada del juego será de $1*(1/6)-1*(5/6)=-4/6$.

Otro ejemplo, bastante más complejo, es cuando se produce un huracán y se quiere estimar cuáles lugares serán afectados, así como la magnitud de los eventuales destrozos. Para ello es necesario construir modelos matemáticos que usen la probabilidad para asegurar que con cierta posibilidad se dirigirá hacia el oeste y tocará tierra en X días, teniendo Y magnitud. Los modelos matemáticos que se requieren para estos cálculos son muy complejos e incluyen las leyes de probabilidad para describir los cambios en el tiempo de la dirección del viento y de la temperatura, entre muchos otros factores.

Distribución de probabilidades

Una distribución de probabilidad indica toda la gama de valores que pueden representarse como resultado de un experimento si éste se llevase a cabo.

Es decir, describe la probabilidad de que un evento se realice en el futuro, constituye una herramienta fundamental para la prospectiva, puesto que se puede diseñar un escenario de acontecimientos futuros considerando las tendencias actuales de diversos fenómenos naturales.

En el orden de las consideraciones anteriores una distribución de probabilidad es un modelo teórico que trata de explicar el comportamiento de un fenómeno real. Actúa como una función que asigna a cada suceso, cuantificado mediante una variable aleatoria, la probabilidad correspondiente; por lo que toda

distribución de probabilidad es generada por una variable (por que puede tomar diferentes valores) aleatoria x (porque el valor tomado es totalmente al azar), y puede ser de dos tipos :

Variable aleatoria discreta (x)

Porque solo puede tomar valores enteros y un número finito de ellos.

Por ejemplo: X Variable que nos define el número de alumnos aprobados en la materia de probabilidad en un grupo de 40 alumnos (1, 2 ,3...ó los 40).

Propiedades de una variable aleatoria discreta (x)

- $p(x_i) < 1$ Las probabilidades asociadas a cada uno de los valores que toma x deben ser mayores o iguales a cero y menores o iguales a 1.
- $\sum p(x_i) = 1$ La sumatoria de las probabilidades asociadas a cada uno de los valores que toma x debe ser igual a 1.

Variable aleatoria continua (x)

Porque puede tomar tanto valores enteros como fraccionarios y un número infinito de ellos dentro de un mismo intervalo.

Por ejemplo: x es la Variable que nos define la concentración en gramos de plata de algunas muestras de mineral (14.8 gr, 12.1, 10.0, 42.3, 15.0, 18.4, 19.0, 21.0, 20.8, ..., n)

Propiedades de una variable aleatoria discreta (x)

$p(x)$ Las probabilidades asociadas a cada uno de los valores que toma x deben ser mayores o iguales a cero. Dicho de otra forma, la función de densidad de probabilidad deberá tomar solo valores mayores o iguales a cero.

El área definida bajo la función de densidad de probabilidad deberá ser de 1.

Función de distribución de probabilidades. Sea X una variable aleatoria discreta cuyos valores suponemos ordenados de menor a mayor. Llamaremos función de distribución de la variable X , y escribiremos $F(x)$ a la función:

$$F(x) = p(X \leq x)$$

La función de distribución asocia a cada valor de la variable aleatoria la probabilidad acumulada hasta ese valor.

Distribución de variables discretas

Una distribución discreta describe la probabilidad de ocurrencia de cada valor de una variable aleatoria discreta. Una variable aleatoria discreta es una variable aleatoria que tiene valores contables, tales como una lista de enteros no negativos.

Con una distribución de probabilidad discreta, cada valor posible de la variable aleatoria discreta puede estar asociado con una probabilidad distinta de cero. Por lo tanto, una distribución de probabilidad discreta suele representarse en forma tabular.

Distribución de variables continuas

Una distribución continua describe las probabilidades de los posibles valores de una variable aleatoria continua. Una variable aleatoria continua es una variable aleatoria con un conjunto de valores posibles (conocido como el rango) que es infinito y no se puede contar.

Las probabilidades de las variables aleatorias continuas (X) se definen como el área por debajo de la curva de su PDF. Por lo tanto, solo los rangos de valores pueden tener una probabilidad diferente de cero. La probabilidad de que una variable aleatoria continua equivalga a algún valor siempre es cero¹⁰.

Valor esperado

El valor que se espera obtener de un experimento estadístico se llama el valor esperado. También llamado “esperanza matemática”. También lo llamamos “media” y esta es la palabra que vamos a seguir usando. Si tiramos una moneda 10 veces, esperamos que salga 5 veces “cara” y 5 veces “cruz”. Esperamos obtener este valor porque la probabilidad de que salga “cara” es 0,5, y si lanzamos la moneda 10 veces, obtenemos 5. Por lo tanto, 5 es la media. Para formalizar este particular ejemplo de la media, si p es la probabilidad y n el número de eventos, la media es $a = np$. Esta es la forma de la media cuando se puede expresar la probabilidad por medio de la distribución binomial.

Para formalizar el concepto un poco más, en un experimento con resultados discretos x_i para los cuales la probabilidad es $P(x_i)$, la media estará dada por:

$a = \sum x_i P(x_i)$ Valor esperado de V.A. continuas Suponga que X es una variable aleatoria continua con función de densidad de probabilidad $f(x)$ La media o valor esperado de X , denotada como μ o E , es $\mu = EX = \int_{-\infty}^{\infty} x f(x) dx$

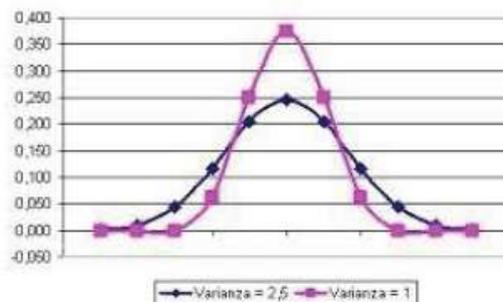
Varianza

La varianza de las variables aleatorias, consiste en una medida vinculada a su dispersión. Se trata de la esperanza del cuadrado de la desviación de esa variable considerada frente su media y se mide en una unidad diferente. Por ejemplo: en los casos en que la variable mide una distancia en kilómetros, su varianza se expresa en kilómetros al cuadrado.

Cabe destacar que las medidas de dispersión (también identificadas con el nombre de medidas de variabilidad) se encargan de expresar la variabilidad de una distribución por medio de un número, en los casos en que las diferentes puntuaciones de la variable están muy alejadas de la media. A mayor valor de la medida de dispersión, mayor variabilidad. En cambio, a menor valor, más homogeneidad.

Lo que hace la varianza es establecer la variabilidad de la variable aleatoria. Es importante tener en cuenta que, en ciertos casos, es preferible emplear otras medidas de dispersión ante las características de las distribuciones.

Se denomina varianza muestral cuando se calcula la varianza de una comunidad, grupo o población en base a una muestra. La covarianza, por otra parte, es la medida de dispersión conjunta de un par de variables.



Los expertos hablan de análisis de la varianza para nombrar a la colección de modelos estadísticos y sus procedimientos asociados en la cual la varianza aparece particionada en distintos componentes.

Desviación estándar

Para muchos la palabra desviación estándar puede sonar desconocida y no la habrán oído nombrar a menos que hayan asistido a una clase de estadística.

Justificadamente la desviación Estándar, en un conjunto de datos (precios en el caso del mercado de valores) es una medida de dispersión, que nos indica cuánto pueden alejarse los valores respecto al promedio (media), por lo tanto es útil para buscar probabilidades de que un evento ocurra, o en el caso del mercado bursátil, determinar entre que rango de precios puede moverse un determinado activo, y determinar qué tipo de activos pueden ser más volátiles que otros.

Los operadores del mercado están interesados en la dirección del precio de un activo y en la velocidad de los movimientos del subyacente para determinar qué tan riesgoso o volátil puede llegar a ser un activo. Los mercados cuyos precios se mueven lentamente son mercados de baja volatilidad, los mercados

cuyos precios se mueven a alta velocidad son mercados de alta volatilidad.

Existen varias maneras de estimar la volatilidad, y el mundo ideal sería aquel donde se pueda determinar la volatilidad de todo el conjunto de datos existentes, sin embargo teniendo en cuenta que se cuentan con recursos (información, costos, etc) limitados, la desviación estándar se puede tomar sobre un determinado conjunto de datos que se ajusten a nuestros requerimientos, mediante la siguiente fórmula:

$$\sqrt{s^2} = \sqrt{\sum_{i=1}^n (x_i - \bar{x})^2 / n - 1}$$

Donde

- xi= dato i que esta entre (0, n)
- x= promedio de los datos
- n= numero datos

Distribución normal

Es la distribución que aparece con mayor frecuencia en el comportamiento de fenómenos reales, en especial en el área de las ciencias naturales. Johann Carl Friedrich Gauss genio matemático, físico y astrónomo, de nacionalidad alemana, fue el que mayormente contribuyó a su formulación y aplicación en diferentes áreas del saber cómo por ejemplo en su aplicación a la teoría de los errores, de importancia en ingeniería.

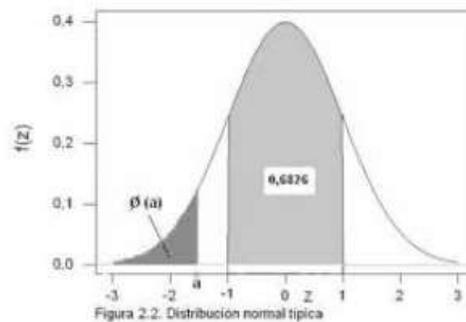
Características de la distribución normal

1. La curva tiene forma acampanada, asintótica al eje hacia $-\infty$ y $+\infty$. El área total encerrada por ésta y es igual a 1, como corresponde a toda función de distribución de probabilidad.
2. La curva tiene un máximo en μ y es simétrica respecto a la recta $x = \mu$. Luego, en estadística son coincidentes la media aritmética, la mediana y la moda, es decir, $\mu = Me = Mo$
3. la curva tiene dos puntos de inflexión que se ubican en $x = \mu - \sigma$ y $x = \mu + \sigma$

4. el área bajo la curva comprendida entre los puntos de inflexión es igual a 0,6826 (68,26%) y el área entre $\mu - 2\sigma$ y 2σ es igual a 0,9544 (95,44%), cualesquiera sean los valores de sus parámetros μ y σ^2 . Se debe recordar que el área bajo la curva, en variables aleatorias continuas, corresponde a la probabilidad de sucesos que son intervalos de números reales. En consecuencia, lo anterior se puede interpretar en el sentido que el 68,26% de los individuos que componen la población teórica tienen un valor de la variable en estudio entre $\mu - \sigma$ y $\mu + \sigma$ y en el 95,44% el valor de la variable quedará comprendida entre $\mu - 2\sigma$ y $\mu + 2\sigma$.

Distribución normal típica

Se llama distribución normal típica a $Z = N(0,1)$, con función de distribución de probabilidad $f(z) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}z^2}$. Su función de distribución acumulativa es $\Phi(z) = \int_{-\infty}^z f(z) dz$ que representa el área bajo la curva normal estándar desde $-\infty$ hasta el valor de real z . Por ejemplo el área acumulada hasta el punto a es $\Phi(a) = \int_{-\infty}^a f(z) dz$.



Se enunciará un teorema de enorme importancia estadística, porque establece la relación entre una distribución normal cualquiera y la distribución normal típica.

Teorema

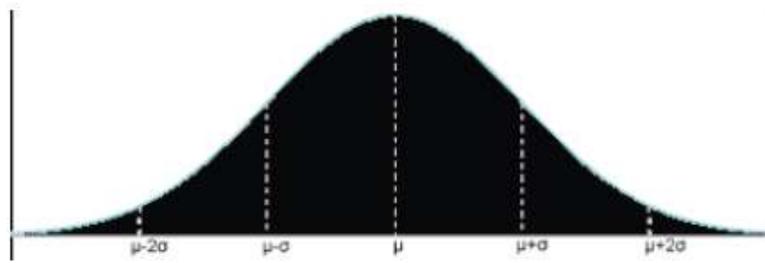
Sea X variable aleatoria con distribución $N(\mu, \sigma^2)$, entonces la variable tipificada $Z = (X - \mu) / \sigma$, tiene distribución $N(0,1)$

Esto es de especial relevancia porque limita los cálculos de probabilidad de distribuciones normales al uso de una tabla única

La distribución normal estándar, o tipificada o reducida, es aquella que tiene por media el valor cero, $\mu = 0$, y por desviación típica la unidad, $\sigma = 1$. Su función de densidad es: Su gráfica es: La probabilidad de la variable X dependerá del área del recinto sombreado en la figura.

Su función de densidad es: $f(x) = \frac{1}{\sqrt{2\pi}} e^{-x^2/2}$

Su grafico es:



Estandarización de la Distribución Normal

Si Z tiene una distribución normal con Esperanza cero y varianza 1, se dice que Z tiene una distribución normal estandarizada. La fdp de Z puede escribirse como:

$$G(z) = \frac{1}{\sqrt{2\pi}} e^{-z^2/2}$$

Teorema

$$\text{Si } X \sim N(\mu, \sigma^2) \text{ e } Y = ax + b \text{ -- } Y \sim N(a\mu + b ; a^2 \sigma^2)$$

La forma estandarizada de cualquier distribución normal se puede obtener aplicando el corolario de este teorema.

$$\text{Si } X \sim N(\mu, \sigma^2) \text{ e } Y = \frac{x-\mu}{\sigma} \text{ -- } Y \sim N(0,1)$$

Propiedad reproductiva de la distribución normal: dadas n variables aleatorias normales e independiente tales que $X_i \approx N(\mu_i, \sigma_i)$, $i=1, \dots, n$, su suma $\sum_{i=1}^n X_i$ sigue también una distribución normal siendo.

BIBLIOGRAFÍA

TEORÍA DE LAS PROBABILIDADES Y DISTRIBUCIÓN DE PROBABILIDADES NORMAL



www.mawil.us

1. Gómez-Torres, E.; Ortiz, J.J. y Gea, M.M. Conceptos y propiedades de probabilidad en los libros de texto españoles de educación primaria. *Avances de Investigación en Educación Matemática*, (2014). n° 5, 49 – 71.
2. Espacio muestral y evento [citado 2 Mar 2019]. Disponible en :<http://navarrof.orgfree.com/Docencia/MatematicasIII/M3UT1/ut1t1.htm>
3. Espacio muestral, eventos y medidas de probabilidad. 2011 [citado 2 Mar 2019]. Disponible en :https://www.cimat.mx/~pepe/cursos/probabilidad_2011/material/material_111109.pdf
4. Acuña E . Concepto basicos de Probabilidades [citado 2 Mar 2019]. Disponible en :<http://academic.uprm.edu/eacuna/miniman4sl.pdf>
5. Sanchez N . Principio aditivo y multiplicativo. 2016 [citado 2 Mar 2019]. Disponible en :<https://prezi.com/bw75f7uwyfln/principio-aditivo-y-multiplicativo>
6. Probabilidad y principio de conteo. 2010 [citado 2 Mar 2019]. Disponible en :<https://jrvargas.files.wordpress.com/2010/07/principio-multiplicativo.pdf>
7. David Ruiz Muñoz y Ana María Sánchez Sánchez. 2006 Apuntes de estadística [citado 2 Mar 2019]. Disponible en :<http://www.eumed.net/libros-gratis/2006a/rms/a6.htm>
8. Distribuciones de probabilidad. Tema 13 [citado 2 Mar 2019]. Disponible en :<https://static1.squarespace.com/static/526e85b4e4b09c47421bd159/t/5876555620099e329ddde427/1484150105476/T13DISPROBb.pdf>
9. Variable aleatoria y función de distribución [citado 2 Mar 2019]. Disponible en :<https://www.ugr.es/~eues/webgrupo/Docencia/MonteroAlonso/estadisticaII/tema2.pdf>
10. Funciones de probabilidad y distribución [citado 2 Mar 2019]. Disponible en :https://www.uv.es/webgid/Descriptiva/3_funciones_de_probabilidad_y_distribucion.html
11. Distribuciones de probabilidad continuas y discretas [citado 2 Mar 2019]. Disponible en :<https://support.minitab.com/es-mx/minitab/18/help-and-how-to/probability-distributions-and-random-data/supporting-topics/basics/continuous-and-discrete-probability-distributions/>
12. Conceptos de estadística aplicada [citado 2 Mar 2019]. Disponible en :<http://hyperphysics.phy-astr.gsu.edu/hbasees/Math/mean.html>
13. Definición de varianza. 2018 [citado 2 Mar 2019]. Disponible en :<https://>

definicion.de/varianza/

14. Mara Laradola .Qué es la desviación estándar y como interpretarla. 2009. [citado 2 Mar 2019]. Disponible en :<https://tradingcenter.wordpress.com/2009/11/11/que-es-ladesviacion-estandar-y-como-interpretarla-1/>
15. Rustom J. A .ESTADÍSTICA DESCRIPTIVA, PROBABILIDAD E INFERENCIA. Una visión conceptual y aplicada. 2012.[citado 2 Mar 2019]. Disponible en :http://repositorio.uchile.cl/bitstream/handle/2250/120284/Rustom_Antonio_Estadistica_descriptiva.pdf?sequence=1

UNIDAD V

INTRODUCCIÓN A LA DISTRIBUCIÓN DE PROBABILIDADES DE VARIABLES CONTINUAS Y DISTRIBUCIONES MUESTRALES



www.mawil.us

Estadísticas inferenciales, objetivos y aplicaciones

La estadística inferencial está concebida sobre la base del cálculo matemático, que tiene como objetivo realizar generalizaciones de determinadas características que aparecen en una muestra previamente calculada y estudiada al resto de la población.

Como se conoce una investigación generalmente no se limita a describir las distribuciones de las variables investigadas, en su pretensión de establecer un comportamiento similar a la muestra en el universo o población, se persigue hacer estimaciones y tomar decisiones que pueden ser de mucha importancia en cualquier esfera de la vida social y económica. Ese procedimiento de generalizar los resultados de la muestra a la población o universo se denomina Inferencia Estadística.

La estadística inferencial puede ser utilizada para dos procedimientos: para comprobar hipótesis y estimar parámetros. La misma se fundamenta en la teoría acerca de la distribución muestral. (Wiersma, 1986, p. 335):

Parámetro y Estimado. El parámetro es un indicador numérico de una población que caracterizan su distribución de probabilidades tales como la proporción, la media, y la varianza (parámetros de la distribución normal) y otros. Estos representan estadísticas de la población. Para conocer acerca de los parámetros de una población no es necesario proceder a recoger los datos de cada uno de sus integrantes, es suficiente con aplicar un estadígrafo que permite hacer el estimado, según la distribución de probabilidad asociada a la variable aleatoria que la representa. Esto se define como Estimación, que es el valor que toma el parámetro en una muestra concreta.

La inferencia de los parámetros se lleva a cabo mediante técnicas estadísticas apropiadas y se selecciona aquel cuya distribución sea conocida y buena para estimar. Aun así, nunca hay una completa seguridad de la estimación efectuada, y aunque el riesgo de cometer error sea mínimo es mejor trabajar con elevados niveles de confianza o seguridad.

Distribución de probabilidad de variables aleatorias continuas

La teoría de la probabilidad se inició como un intento de responder a un grupo de preguntas que surgían en los juegos de azar, por ejemplo, conocer cuántos dados hay que lanzar para que la probabilidad de que salga algún seis supere el 50%. Siendo más preciso en este tipo de ejemplo hay que decir que, al lanzar un dado ideal, la probabilidad de cada una de las caras pueda salir es $1/6$, luego entonces al lanzar dos dados, la probabilidad de cada uno de los resultados es $1/36$. Es obvio que este cálculo de las probabilidades genere la obtención de una determinada muestra.

El cálculo matemático de probabilidades se fundamenta en situaciones teóricas en las cuales puede configurarse un espacio muestral cuyos eventos esenciales alcancen todos igual probabilidad, cuyo resultado se representa con un número entre 0 y 1, ambos inclusive. La probabilidad 0 indica que el resultado no ocurrirá nunca, y la probabilidad 1, que el resultado ocurrirá siempre.

Dentro de los conceptos medulares de la teoría de las probabilidades está el de variable aleatoria que, se define como cualquier característica que adquiere distintos valores con probabilidades determinadas, que cuando se trata de una variable continua puede tomar cualquier valor de un intervalo, y que para esta variable existe la llamada ley normal para variables continuas como parte de las leyes de probabilidad teóricas.

En la distribución de probabilidad de variables aleatorias continuas se encuentra: la distribución normal (es la más importante del Cálculo de probabilidades y de la Estadística), la t de Student, el Chi Cuadrado y la varianza F (igualmente importantes).

Una variable estadística es continua si admite todos los valores de un intervalo, como ocurre con la estatura.

Distribución normal: Características, obtención de probabilidades. Propiedades de simetría. La distribución normal de una muestra (Curva normal) queda completamente definida a través de dos parámetros: la media (μ) y la desvia-

ción estándar (Sigma), también es denominada curva o campana de Gauss, en honor al matemático alemán Carl Friedrich Gauss, es la distribución media o promedio de las características de una población, cuya gráfica produce una figura tipo acampanada.

Dentro de sus características es posible señalar que es una distribución continua de frecuencia, tiene aplicación usual en cualquier situación y posee un carácter simétrico. El cálculo de la media \bar{X} es el resultado de la división de la suma de todos los valores observados entre el número de observaciones. Su fórmula es:

$$\bar{X} = (X_1 + X_2 + \dots + X_n) / n$$

Si se tienen 7 niños que pesan en Kilogramos: 8, 7, 10, 6, 7, 11, 9

$$\bar{X} = (8 + 7 + 10 + 6 + 7 + 11 + 9) / 7 = 8,3 \text{ Kg}$$

Como se puede observar el promedio aritmético o media tiene la ventaja de tomar en cuenta la totalidad de los valores de la serie, pero por esta razón se puede afectar desventajosamente por valores que sean anormalmente altos o bajos.

En cuanto a la desviación estándar, esta se puede calcular por la fórmula:

$$\sigma^2 = \sum_{i=1}^N (x_i - \bar{x})^2 / n$$

También se puede emplear esta otra fórmula:

$$\sigma^2 = (\sum f \cdot d^2) / n$$

Donde la desviación estándar resulta de la sumatoria de las multiplicaciones obtenidas del número de individuos de cada intervalo por las desviaciones elevadas al cuadrado, divididas por el total de la muestra estudiada.

- 5.2 Estandarización de variables aleatorias normales. Aplicaciones. Propiedad Reproductiva. Generación de probabilidad t-de Student. Generación de la Distribución de Probabilidad Chi-Cuadrado. Lectura de tablas de probabilidad.

Estandarización de variables aleatorias normales. Aplicaciones

La distribución de muchas variables, como los caracteres morfológicos de individuos —altura, peso o longevidad—, caracteres fisiológicos, sociológicos, psicológicos o físicos y, en general, cualquier característica que se obtenga como suma de muchos factores, sigue la curva normal.

Los valores de una población normal (μ , σ) pueden transformarse en valores de una población estandarizada. Este proceso de normalización transforma la función original $N(\mu, \sigma)$ en una función estandarizada $N(0, 1)$, resultando interesante en la práctica, ya que de esta distribución existen tablas publicadas a partir de las que se puede obtener la probabilidad de observar un dato menor o igual a un cierto valor z . La ecuación de estandarización es:

$$Z = (X - \mu) / \sigma$$

Propiedad reproductiva

Generación de probabilidad t-de Student

La prueba t de Student es un método estadístico paramétrico, que parte de considerar la distribución de la variable dependiente es normal en el universo o población, y su medición se realiza por intervalos o razón.

Es una prueba estadística que permite evaluar si existe diferencia significativa entre las medias de dos grupos. Para ello se establecen dos hipótesis: de investigación y nulidad; la primera plantea la diferencia significativa entre ambos grupos con relación a una determinada variable y la hipótesis nula, niega la existencia de tal diferencia.

La prueba t de Student se puede realizar entre dos grupos para una o más variables, pero siempre se realiza de manera independiente. Se aplica a muestras grandes con la fórmula:

$$t = (X_1 - X_2) / \sqrt{S_1^2/N_1 + S_2^2/N_2}$$

Donde:

X_1 es la media del primer grupo, que es determinado por el investigador, en tanto

X_2 es la media del segundo grupo, objeto de la correlación.

S_1^2 es la desviación estándar del primer grupo elevada al cuadrado, N_1 es el tamaño del primer grupo.

S_2^2 ; es la desviación estándar del segundo grupo elevada al cuadrado y N_2 es el tamaño del segundo grupo.

Un elemento esencial en el resultado de esta prueba es el cálculo de los grados de libertad, que se calcula de una forma muy simple, se suman ambas muestras y se le resta dos unidades:

$$gl = (N_1 + N_2) - 2$$

Teniendo en cuenta los valores obtenidos en “t” y los grados de libertad, se busca en la tabla de distribución de “t” de Student, el nivel de significancia que puede existir en la correlación de ambas muestras. En el caso de que el valor calculado sea igual o mayor al que aparece en la tabla, se acepta la hipótesis de investigación, en caso contrario (sea menor), se confirma la hipótesis nula, todo ello tomando en consideración el nivel de confianza elegido (0.05 o 0.01) y los grados de libertad. Cuando más alta es la cifra de grados de libertad se aproxima más a la normalidad la distribución de esta prueba. (Wiersma, 1986).

Generación de la Distribución de Probabilidad Chi-Cuadrado. Lectura de tablas de probabilidad. El Chi cuadrado o Ji cuadrada (X^2) es una prueba estadística no paramétrica que se emplea para evaluar hipótesis acerca de la relación entre dos variables categóricas, la misma no considera relaciones causales.

Con esta prueba se miden variables nominales u ordinales, también pueden ser de intervalos o razón.² Con el X² se pretende comparar el comportamiento que existe entre dos variables para comprobar si la relación entre ellas es o no significativa, de acuerdo con los grados de libertad obtenidos y el porcentaje de significancia de un 95% o un 99%.

Para la determinación del Chi cuadrado se requiere de una tabla de contingencia o tabulación cruzada donde aparecen encabezándola las dos variables que se desean medir, las que a su vez se subdividen en dos o más categorías. Una variable podría ser perfumes y la evaluación recibida por un grupo integrado. Vamos a denominar a los perfumes A, B y C; y en el nivel de ventas semanales tendremos: Primera, segunda, tercera y cuarta semana.

Tabla 1.

Perfumes	Niveles de venta		Total
	Boutique Norte	Boutique Sur	
Perfume A	70	90	160
Perfume B	85	65	150
Perfume C	105	95	200
Total	260	250	510

Cuando contamos con los datos de la tabla anterior procedemos de la manera siguiente:

1. Obtener las frecuencias esperadas o teóricas (FE). Se realiza multiplicando el total de cada columna por el subtotal de cada fila, dividido por el total general, ejemplo: $260 \times 160 / 510 = 81,6$; $260 \times 150 / 510 = 76,5$; y así sucesivamente hasta finalizar con esos cálculos y colocarlos en una tabla para su mejor manejo.

Tabla 2

81,6	78,4	160
76,5	73,5	150
102,0	98,0	200
260	250	510

2. Se resta cada valor observado (FO) con su correspondiente valor esperado (FE).
3. La diferencia resultante (O-E) se elevará al cuadrado y se dividirá por su respectiva frecuencia esperada. (Ver Tabla 3)

Tabla 3

Perfumes	Ventas Boutique Norte				Ventas Boutique Sur			
	FO	FE	(O-E)	(O-E) ² /E	FO	FE	(O-E)	(O-E) ² /E
A	70	81,6	11,6	1.64	90	78,4	11,6	1.71
B	85	76,5	-8,5	0.94	65	73,5	-8,5	0.98
C	105	102	-3	0.08	95	98,0	-3	0.09

4. Se suman todos los resultados obtenidos de la formula $(O-E)^2/E$, que será el valor de Chi cuadrado, o sea:

$$X^2 = 1,64+0,94+0.08+1.71+0.98+0,09=5.44$$

Para poder interpretar el valor de Chi Cuadrado (X^2) es necesario conocer su grado de libertad. El grado de libertad de X^2 se realiza multiplicando el número de columnas menos uno por el número de renglones menos uno:

$(c-1) (r-1)$, donde en nuestro caso, hay 3 renglones correspondiente a los 3 perfumes y 2 columnas correspondientes a las 2 boutique. De acuerdo con esto, el grado de libertad sería: $Gl = (3-1) (2-1) = 2 \times 1 = 2$.

En la distribución de la Tabla de Chi cuadrado para dos grados de libertad el valor para un 95% de certeza es de 5.991 y para el 99% de certeza es de 9.210.

Como se puede apreciar el resultado de X^2 obtenido (5.44) es inferior por lo tanto la conclusión es que las diferencias observadas se pueden explicar fácilmente por el azar y no es necesario recurrir a otras explicaciones.

Para haber considerado la existencia de alguna correlación el valor hallado de Chi cuadrado debía haber sido igual o superior al que aparece en la tabla de distribución para esta prueba con 2 grados de libertad.

Generación de la Distribución de Probabilidad Fischer (F)

La prueba estadística paramétrica F o de varianza permite analizar si en más de dos grupos aparecen diferencias significativas en cuanto a sus medias y al valor de las varianzas.

Para la realización de esta prueba se plantea como hipótesis de investigación que los grupos difieren significativamente entre sí y la hipótesis nula establece que entre los grupos no hay diferencia significativa.

Esta distribución de probabilidad está caracterizada por ser completamente asimétrica y tener dependencia de dos parámetros o grados de libertad. Esto quiere decir que si de dos poblaciones normales, o cercanamente normales, se escogen dos muestras aleatorias e independientes, y a cada una se le calcula su respectiva varianza, el cociente de ambos valores F tendrá una distribución de Fisher, cuyos valores críticos fueron obtenidos por W. Snedecor.

Esta tabla se caracteriza por tener dos grados de libertad: el correspondiente al numerador $n_1 - 1$ y el del denominador $n_2 - 1$. Esto se define matemáticamente de la siguiente manera:

$$F_{n,m} = (x_n^2/n)/(x_m^2/m)$$

La distribución de “F” es asimétrica y comienza del valor “0”, no posee valores negativos, al igual que la distribución X^2 .

Lectura de tablas de probabilidad. Uso de software. Existen tablas que per-

mite interpretar los resultados que se obtienen de la aplicación de las diferentes pruebas estadísticas que son empleadas en las diversas investigaciones científicas.

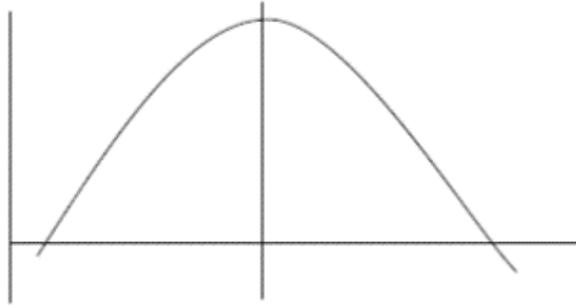
En la actualidad el procesamiento de los datos obtenidos en las investigaciones se realiza a través de diferentes programas de computación que agilizan el trabajo de los investigadores.

5.4. Distribuciones Muestrales. Definición y propiedades. Teorema central del límite. Distribución de probabilidad de Media Muestral cuando $n > 30$ y cuando $n \leq 30$. Distribución de probabilidad de la proporción muestral. Aplicaciones. Distribución de la probabilidad de la Varianza.

Distribuciones Muestrales. Definición y propiedades
Teorema central del límite

Las distribuciones muestrales son aquellas que se producen de manera tal que en sus resultados, la mayoría de los valores están muy cerca del verdadero valor de la población estudiada. Estos resultados no son desordenados, sino que presentan un grado de simetría. Las frecuencias de la distribución aumentan de forma paulatina y luego decrecen de la misma manera. De esta agrupación puede observarse la formación de una curva en forma de campana. lo óptimo de una muestra depende de cuánto se aproxima su distribución a la distribución de las características de la población.

Más abajo se hace una representación de la forma que adquieren los valores en la curva.



Es característico que la aproximación de la muestra a la distribución de las características de la población mejore al incrementarse su tamaño. La distribución normal es la que toma forma de campana y se alcanza generalmente con muestras de 100 o superiores a esta cifra. Esto resulta muy útil cuando existe la necesidad de hacer inferencias estadísticas.

Esta tendencia que se genera en la muestra de una distribución normal se conoce con el nombre de teorema del límite central y es lo que posibilita estimar los valores de la población, de modo inferencial.

El teorema central del límite es uno de los resultados fundamentales de la estadística porque establece que cuando una muestra es lo suficientemente grande, es decir es mayor de 30 sujetos, sea cual sea la distribución de la media muestral, esta se mantendrá cercana a una distribución normal, por lo tanto en cualquier variable aleatoria, si extraemos muestras de tamaño n ($n > 30$) y calculamos sus promedios, estos promedios continuarán una distribución normal. Igualmente, la media será equivalente a la de la variable de interés, y la desviación estándar de la media muestral será aproximadamente el error estándar.

Un ejemplo específico del teorema central del límite es la distribución binomial. A partir de $n=30$, la distribución binomial tiene un comportamiento estadístico como una curva normal, por lo que es posible aplicar los tests estadísticos adecuados a esta distribución.

La importancia del conocimiento y dominio del teorema central del límite está en que mediante un conjunto de teoremas, se revelan los fundamentos por

los que estas distribuciones normales o casi normales aparecen en todo momento en diferentes campos de aplicación.

Distribución de probabilidad de Media Muestral cuando $n \geq 30$ y cuando $n \leq 30$.

En cierta medida con anterioridad se ha puntualizado que si el tamaño de la muestra es suficientemente grande ($n \geq 30$) entonces, para casi todas las poblaciones, la media muestral \bar{X} sigue aproximadamente una distribución normal pero en el caso de que $n < 30$, la aproximación sería confiable solo si el comportamiento de la población no difiere mucho de una distribución normal y además se conoce que la población es normal. En este caso, sin importar cuán reducida sea muestra, la distribución seguirá una curva normal exacta.

Recordemos que la media muestral \bar{X} obtenida de una muestra aleatoria de tamaño n de una población con media μ y varianza σ^2 , tiene una distribución normal con media μ y varianza σ^2/n .

Vamos a poder medir qué tanto se desvía la media muestral de la media poblacional a través del valor Z , de la siguiente manera:

$$Z = \frac{\bar{X} - \mu}{\sigma_{\bar{X}}} = \frac{\bar{X} - \mu}{(\sigma/\sqrt{n})} = \frac{(\bar{X} - \mu)\sqrt{n}}{\sigma}$$

Es fácil ver que la Z , que es una estandarización de la media muestral, sigue una distribución $N(0,1)$, que cuando $n \geq 30$ y cuando $n \leq 30$.

Distribución de probabilidad de la proporción muestral. Aplicaciones.

La proporción muestral es aquella parte de una población de la que se extraen muestras de tamaño $n \geq 30$, y de la que se sabe que la proporción de individuos que tienen una cualidad específica es igual a p . La variable aleatoria \hat{p} de las proporciones muestrales es la proporción de individuos de cada muestra que presentan la característica estudiada. Se define como $\hat{p} = X/n$, donde X es el número de éxitos y n el tamaño de la muestra.

Se tiene que \hat{p} sigue una distribución normal

$$N(p, \sqrt{p(1-p)/n})$$

La varianza σ^2 es desconocida pero n es grande. Cuando la varianza no es conocida, la distribución de la media \bar{X} es una t-Student, que para tamaño muestral $n \geq 30$ se puede aproximar por una $N(0, 1)$.

Distribución de la probabilidad de la Varianza

En muchas ocasiones no conocemos la probabilidad de éxito en un experimento binomial y tiene que ser estimado de la muestra. Como p es la probabilidad de éxitos en cualquier prueba, en una población finita, p mide la proporción de éxitos en esa población.

Así, si en una muestra de tamaño n de una población, X es el número de éxitos, estimamos la proporción de éxitos en esta muestra: x/n

Entonces $\hat{p} = x/n$ tiene una distribución normal con media p y varianza $p(1-p)/n$ siempre y cuando $np(1-p) > 5$

Muchos problemas están enfocados en determinar si la proporción de gente o cosas en una población que posee cierta característica es la misma que la proporción que posee dicha característica en otra población: $p_1 = p_2$, ó si es mayor: $p_1 > p_2$ ó menor: $p_1 < p_2$.

Cuando desconocemos estas proporciones es necesario tomar una muestra de cada población y estimar dichas proporciones.

UNIDAD VI

TÉCNICAS DE MUESTREO Y ESTIMACIÓN DE PARÁMETROS POBLACIONALES



www.mawil.us

Muestreo Estadístico

Antes de incursionar en el muestreo estadístico y para evitar cualquier tipo de confusión, es necesario dominar dos conceptos claves: universo o población y muestra.

El universo hace referencia al conjunto de todos los elementos que son objeto de una investigación, y la muestra es un subconjunto del universo, que se selecciona de manera proporcional para poderlo representar.

Según el criterio de Bisquerra (1989), el muestreo es el procedimiento para seleccionar la muestra de individuos sobre los que se van a recoger los datos y que deben ser representativos de la población que se desea estudiar. Esto es un requisito fundamental para poder llegar a conclusiones que sean generalizables.

La muestra incluye un conjunto de personas o cosas bien determinadas que reúnen determinados requisitos, de acuerdo con los fines de una investigación. Esta muestra es escogida entre los integrantes de la población o universo porque en muchas oportunidades es imposible estudiar la población en su totalidad por ser demasiado grande, y por tanto el investigador debe precisar las características de los sujetos o cosas que desea estudiar para que toda la población o universo se vea representada en ella. Ej. El número o porcentaje de adolescentes escolarizados que son fumadores. La población o universo son todos los adolescentes que estudian en centros docentes. De esa población que muy grande es necesario seleccionar una muestra representativa que permita realizar conclusiones a partir de los hallazgos encontrados, a toda esa población.

Cuando el universo o población es un conjunto pequeño, es posible abarcarlo en su totalidad porque no requiere de muchos esfuerzos ni recursos para realizar la investigación. Ejemplo: El consumo de café en los estudiantes que asisten a las olimpiadas de historia.

De este modo podemos precisar que la muestra es una parte o subconjunto del universo, que para su selección existen criterios y procedimientos estadísticos que deben ser seguidos por el investigador para errores en la interpretación

del fenómeno que se desea estudiar.

Cuando se trata de un estudio exploratorio, la muestra se escoge fundamentalmente por su tamaño, sin embargo en la investigación descriptiva o experimental, la muestra debe ser representativa, aunque sin dejar de tomar en cuenta su tamaño.

En muchas oportunidades, cuando se desea garantizar la representatividad de la muestra, se trabaja con una muestra piloto y de acuerdo con los resultados obtenidos se determina el tamaño de la muestra definitiva.

Métodos de muestreo

Existen diferentes tipos de muestreo entre los que se encuentran: el probabilístico o aleatorio y el no probabilístico o intencional.

Muestreo Probabilístico

Los métodos de muestreo probabilísticos están basados en el principio de equiprobabilidad, lo que representa que para todos los individuos de la población existen iguales probabilidades de integrar la muestra.

Se recomienda utilizar métodos probabilísticos siempre que sea posible, puesto que aseguran mejor la representatividad de la muestra. Los métodos probabilísticos se clasifican en:

Muestreo Aleatorio Simple

El procedimiento utilizado consiste en asignar un número a cada individuo de la población y hacer la selección mediante algún sistema mecánico, que puede ser a partir de tablas de números aleatorios, sorteo u otro procedimiento similar. En este tipo de muestreo lo importante no es el sujeto sino la característica que se quiere conocer.

Muestreo Aleatorio Estratificado

Se utiliza cuando además de tener la misma posibilidad de ser escogidos, los elementos muestrales o unidades de análisis tengan un determinado atributo que sea del interés de la investigación, que pudiera ser el sexo, la edad, la profesión, la calificación ocupacional o cualquier otro. Para ello se divide la población en los estratos seleccionados, y dentro de cada estrato se realiza un muestreo.

Muestreo por conglomerados

En éste en vez de escoger individuos, se seleccionan conglomerados de personas pertenecientes a escuelas, hospitales, industrias, localidades, municipios, etc. Estos conglomerados que son escogidos de manera aleatoria, son tomados en su conjunto para estudiar cada uno de los integrantes. Si se desea investigar acerca del desarrollo de la economía social y solidaria en un país y su impacto en la sociedad, se pueden escoger territorios donde estén enclavadas esa forma de economía. Cuando la población sea muy grande y no se pueda abarcar en su totalidad, luego de seleccionar los conglomerados se estratifican y se determina la muestra dentro de cada estrato.

Muestreo Bietápico y Trietápico

En algunas oportunidades cuando se quiere extraer una muestra de la Población o Universo existe el inconveniente de no contar con un listado de toda esa población o que la misma está distribuida en un territorio muy extenso. Esto obliga al investigador a realizar el muestreo en dos o tres etapas, y aun en más.³

Por ejemplo si se desea realizar una encuesta en la población estudiantil universitaria de una provincia, se emplearía un procedimiento trietápico porque se obtendría la muestra en tres etapas, ello consistiría en:

1. Listar las universidades y seleccionar por sorteo un número de ellas.
2. Escoger al azar un determinado número de Facultades en cada Universidad escogida.
3. Determinar de manera aleatoria el número de estudiantes a encuestar.

Sería bietápica en caso que el estudio se decidiera hacer en una determinada universidad, donde solo se efectuarían las etapas 2 y 3.

Selección del tamaño de una muestra completamente aleatoria

La selección de la muestra constituye un momento crucial para la realización de una investigación que cumpla con un elevado grado de objetividad porque la variable que se desea estudiar en una población sea lo suficientemente representativa en esa muestra.

Autores como Bugeda (1975) y Fox (1981), ofrecen un conjunto de criterios y tablas que orientan en cuanto a una selección de una muestra, teniendo en cuenta su tamaño y nivel de confianza (respecto a la representatividad con relación a la población).

De acuerdo con lo anterior se plantean fórmulas, que responden a los criterios de selección de una muestra infinita (población de más de 100 000 individuos), o de selección de una muestra finita (población de menos de 100 000 individuos).

- 1a. Cuando la población es infinita, en la variante, que por estudios previos se conozca la proporción de individuos con la característica que se investiga dentro de la población se plantea la siguiente fórmula:

$$n = (Z_{\infty}^2 \cdot p \cdot q) / e^2$$

n = número de elementos que deben integrar la muestra.

∞ = nivel de significación.

Z_{∞} = puntuación que corresponde al riesgo ∞ elegido.

Ejemplo, para un riesgo $\infty = 0,05$ ($Z_{\infty=1,96}$)

p = % estimado.

q = 100-p

E = error permitido

Ejercicio

Calculemos la muestra para conocer la opinión de una población con respecto a la calidad de un producto, teniendo como datos: nivel de confianza del 99,7(3 σ) y un error de estimación del 5%, conociendo que p=25%

$$n = (3^2 \cdot 25 \cdot 75) / 5^2 = 675 \text{ personas}$$

- 1b. Cuando se desconoce la proporción de la muestra en la población, se dan valores a p y a q del 50%, y se puede calcular la muestra con un nivel de confianza del 99,7(3 σ) y un error de estimación admitido del 2%. Esto se representa en la formula siguiente:

$$n = (3^2 \cdot 50 \cdot 50) / 2^2 = 5625 \text{ personas}$$

Si es necesario estimar alguna media, se aplica la fórmula:

$$n = (Z_{\infty}^2 \cdot \sigma^2) / e^2$$

- En aquellos casos, en los que la población es finita, o lo que es lo mismo no sobrepasa los 100000 sujetos, se procede de la manera que sigue:

Se desea conocer cuántos sujetos deben integrar una muestra para que sea representativa en una población de 20 000 vendedores, con un nivel de confianza de 2 σ (95,5%) y un margen de error del 3%. Se conoce que el 20 % son vendedores ambulantes. La fórmula a aplicar es:

$$n = (Z_{\infty}^2 \cdot p \cdot q \cdot N) / (e^2 (N-1) + Z_{\infty}^2 \cdot p \cdot q)$$

$$n = (4 \cdot 20 \cdot 80 \cdot 20\ 000) / (9(20000-1) + 4 \cdot 20 \cdot 80) = 687 \text{ vendedores}$$

Si es necesario estimar alguna media, se aplica la fórmula:

$$n = (N \cdot Z_{\infty}^2 \cdot \sigma^2) / (Z_{\infty}^2 \cdot \sigma^2 + e^2 \cdot (N - 1))$$

Muestreo No probabilístico o dirigido

Como lo indica su nombre, en este tipo de muestreo la elección de los sujetos depende de la decisión del investigador y no de que todos tengan la misma probabilidad de ser elegidos, por tanto es fundamental el conocimiento y su experiencia investigativa. Con este muestreo dirigido, las pruebas estadísticas tienen un valor limitado y relativo a la muestra, pero no así con relación a la población. No obstante estos métodos seleccionan a los integrantes de la muestra teniendo en cuenta determinados criterios, que garanticen que la muestra escogida sea lo más representativa posible, y realmente son útiles en determinados estudios, donde la representatividad de la muestra con relación al universo no constituye lo esencial, y si lo son las características asociadas al problema que se investiga.⁴ Los principales son:

Muestreo por conveniencia

En esta categoría se comprenden todas las muestras que son seleccionadas por la opinión de un experto, que identifica aquellos representantes típicos del universo o población, basado en su experiencia. Existe una gran flexibilidad en el momento de escoger los individuos que integrarán la muestra, lo cual facilita el trabajo, sin embargo, la validez de los resultados que se obtengan estarán en dependencia del nivel de acierto que se haya tenido al seleccionar esa muestra. Es evidente que no se puede hablar ni de la precisión, ni del carácter fidedigno de los resultados que se alcanza en un probabilístico.

Este muestreo responde en ocasiones a limitaciones existentes en cuanto a recursos, a la imposibilidad de obtener la lista completa de la población a estudiar o a la necesidad de localizar a los individuos deseados en una población muy numerosa.

Muestreo a juicio de expertos

En este tipo de muestreo se trata de que la estructura de la muestra coincida con la de la población de ser conocida. Tiene cierta similitud al muestreo estratificado porque para obtener la muestra se procede a: clasificar la población que

se conozca (sectores, ramas, entidades), decidir el tamaño óptimo de la misma y seleccionar los elementos que la integrarán.

Los valores muestrales obtenidos sirven para la realización del estudio, sin embargo no deben ser utilizados para realizar inferencias.

Muestreo por cuotas

Este tipo de muestra se emplea bastante en estudios de mercadotecnia y para conocer estados de opinión pública. Los encuestadores son instruidos en la manera de aplicar los cuestionarios a las personas que se encuentran en la calle, así van llenando cuotas de acuerdo con la proporción de ciertas variables demográficas en la población. De esta forma ocurre cuando se desea conocer la aceptación de un nuevo producto o servicio por la población a la cual va dirigido. En estos casos, una vez definida la población (universo) se establecen cuotas muestrales que sean representativas. Si es un producto para el uso de la tercera edad sin distinción del sexo, se distribuye el 50% para los hombres y el 50% para las mujeres. Es evidente que el criterio del investigador es importante para definir como se establecen y distribuyen las cuotas (que incluye la homogeneidad máxima posible de los individuos que la integran).

Muestreo tipo Bola de Nieve

El muestreo de bola de nieve tiene una condición no probabilística y se emplea cuando existen dificultades para hallar los participantes potenciales o cuando la muestra se limita a un subgrupo muy pequeño de la población.

El reclutamiento por parte de los participantes a otros conocidos por ellos con características similares constituye una regularidad que contribuye a la investigación. El juicio que se emplea en la selección de la muestra está a cargo de los investigadores. El muestreo de bola de nieve tiene dos pasos: el primero es identificar sujetos potenciales en la población y el segundo, solicitarle su ayuda para reclutar a otros sujetos que puedan integrar la muestra, esto manteniendo la ética profesional. Por esa razón recibe este muestreo el nombre “bola de nieve”.

- 6.2. Estimación de los parámetros poblacionales. Estimación puntual. Estimación interválica: Definición y aplicaciones. Las propiedades de los buenos estimadores. Estimación puntual de la media cuando $n > 30$ y cuando $n \leq 30$, Proporción y Varianza Poblacional, Obtención del tamaño de la muestra para el caso de la estimación de la Media y para el caso de la estimación de la proporción, cuando N es conocido y desconocido.

Estimación de los parámetros poblacionales

El objetivo más importante de la Estadística es obtener una inferencia con respecto a la población basándose en la información contenida en una muestra. Como las poblaciones se describen mediante medidas numéricas denominadas parámetros, el objetivo de la mayoría de las investigaciones estadísticas es deducir una inferencia con respecto a uno o más parámetros de la población.

Los problemas que se tratan en la inferencia estadística se dividen generalmente en dos clases: los problemas de estimación y los de prueba de hipótesis. Como al estimar un parámetro poblacional desconocido se suele hacer una afirmación o juicio este último ofrece solamente una estimación. Es un valor particular obtenido de observaciones de la muestra.

Un parámetro es una medida numérica de algún aspecto de la población. Cuando no tenemos la información sobre toda la población es necesario estimar el valor del parámetro en base a la información de la muestra sobre dicho aspecto de interés y tenemos lo que se llama “estadística” (V.E.Rohen)

Un parámetro describe una población de la misma manera que una estadística describe a una muestra.

Estimación puntual

Una estimación es puntual cuando se usa un solo valor extraído de la muestra para estimar el parámetro desconocido de la población. Al valor usado se le llama estimador. Es posible hacer la estimación puntual de la media de la población a través de la media muestral, de la misma forma la proporción de

la población se estima puntualmente mediante la proporción de la muestra, así como la desviación típica de la población se consigue estimar puntualmente a partir de la desviación típica de la muestra.

Estimación por intervalos

En oportunidades es aconsejable obtener unos límites entre los cuales se ubique el parámetro con un determinado nivel de confianza, en este caso se hace referencia a la estimación por intervalos. Las propiedades de los buenos estimadores.

Los estimadores son considerados como buenos cuando con sus resultados se pueden hacer inferencias de la muestra a la población con un elevado nivel de objetividad, para ello tienen que cumplir con algunas propiedades como son:

1. Insesgabilidad. Existe cuando la esperanza del estimador es igual al valor del parámetro que se desea estimar o sea en este caso el sesgo es nulo.
2. Eficiencia. si se utilizan dos estadísticos como estimadores del mismo parámetro, entonces aquel cuya distribución muestral tenga menor variancia, es un estimador más eficiente o más eficaz que el otro.
3. Consistencia: Dado el caso que no exista la posibilidad de utilizar estimadores de mínima varianza, entonces la consistencia del estimador radica en que a medida que el tamaño de la muestra crezca el valor del estimador tenga correspondencia con el valor del parámetro poblacional. En símbolos se representa así:

$$\lim_{n \rightarrow \infty} E(\hat{\theta}) = \theta \text{ y } \lim_{n \rightarrow \infty} V(\hat{\theta}) = 0$$

1. Suficiencia: Un estimador suficiente del parámetro es aquel no tiene pérdida de la información, θ brinda una información tan exacta como la que se puede disponer en la muestra.

Estimación puntual de la media cuando $n > 30$ y cuando $n \leq 30$, proporción (p) y Varianza (σ^2) Poblacional. Como ya se ha explicado, la estimación puntual es

un procedimiento de la estadística inferencial mediante el cual se realizan cálculos con los datos de una muestra cuyo resultado es un valor numérico único empleado para estimar el valor de un parámetro poblacional. Este procedimiento se hace empleando estimadores, que representan la columna vertebral de la inferencia estadística, los mismos se utilizan en muchos casos prácticos.

El estimador puntual de la media muestral es

$$\bar{X} = (\sum X_i)/n$$

Y el parámetro poblacional que se desea investigar es $\mu = (\sum X_i)/N$, que constituye la media poblacional. En el caso del estimador de la varianza muestral tenemos $S^2 = (\sum (X - \bar{X})^2)/(n-1)$, que permite el cálculo de la varianza poblacional, representado por:

$$\sigma^2 = \sum (X - \mu)^2/N$$

La estimación puntual de la proporción muestral está dada por:

$$\hat{p} = (\sum_{i=0}^n X_i)/n$$

La proporción de la población se puede estimar puntualmente mediante la proporción de la muestra, teniendo en cuenta que para una n suficientemente grande, la distribución es normal.

Estimación interválica de la diferencia de medias. Aplicaciones

Como se mencionó previamente, cuando se trabaja con muestras grandes la desviación estándar de la población es muy similar a la desviación estándar de la muestra y el teorema central del límite garantiza que la distribución muestral de la media sea normal. En cambio, si se tienen muestras pequeñas y se desconoce la desviación estándar poblacional se puede acudir al auxilio de la distribución t student, siempre y cuando se conozca que la población tiene una distribución normal.

Existen casos en los que es necesario estimar la diferencia entre dos medias, con la finalidad de comparar dos poblaciones, por ejemplo:

El estimador puntual de la diferencia entre μ_1 y μ_2 , lo da el estimador $X_1 - X_2$. Por lo tanto, para obtener una estimación puntual de μ_1 y μ_2 se seleccionarán dos muestras aleatorias independientes, una para cada población, de tamaños n_1 y n_2 , y se calculará la diferencia entre sus medias muestrales.

En el caso de trabajar con muestras grandes de cualquier tipo o que se conozca que la población tiene una distribución normal y la desviación estándar poblacional sea conocida, la normal estandarizada estaría dada por:

Estimador $T = X - Y$

La distribución muestral

$$Z = (x - y) - (\mu_x - \mu_y) / \sqrt{((\sigma_x^2)/n_x + (\sigma_y^2)/n_y)} \rightarrow N(0; 1)$$

El intervalo de confianza correspondiente estará comprendido entre $-Z_{\infty/2}$ y $Z_{\infty/2}$, sustituyendo en la fórmula de la normal estandarizada se tiene:

$$-Z_{\infty/2} = (x - y) - (\mu_x - \mu_y) / \sqrt{((\sigma_x^2)/n_x + (\sigma_y^2)/n_y)} = Z_{\infty/2}$$

Estimación de la diferencia entre dos proporciones poblacionales.

Frecuentemente se presentan casos donde es necesario tomar decisiones a partir de la estimación de dos proporciones. En este caso, la finalidad de la estimación consiste en calcular las diferencias o similitudes que existen entre dos proporciones de poblaciones diferentes.

Esta situación se presenta en muchos casos relacionados con los negocios o las ciencias sociales.

En este tipo de casos es importante contar con un medio que permita estimar la diferencia que existe entre las proporciones de dos poblaciones y decidir de qué manera hemos de llevar a cabo el análisis e interpretación de sus resultados. Un procedimiento que facilita esta labor es la estimación de la diferencia entre

proporciones a través de intervalos de confianza. Este procedimiento se puede aplicar a partir de elegir dos muestras independientes n_1 y n_2 de dos poblaciones binomiales, si X_1 y X_2 son los números de aciertos o éxitos que se obtienen al muestrear n_1 y n_2 , entonces se pueden formar las proporciones.

$$\hat{p}_1 = X_1/n_1 \text{ y } \hat{p}_2 = X_2 / n_2$$

El estimador puntual de la diferencia de proporciones de dos poblaciones.

$\hat{p}_1 - \hat{p}_2$ es $\hat{p}_1 - \hat{p}_2$ Considerando la distribución muestral de $\hat{p}_1 - \hat{q}_2$, puede construirse un intervalo de confianza para estimar $\hat{p}_1 - \hat{p}_2$.

Si se tienen muestras lo suficientemente grandes de tal manera que $n_1 \hat{p}_1, n_1 \hat{q}_1, n_2 \hat{p}_2, n_2 \hat{q}_2$, son mayores a 5, la distribución muestral de $\hat{p}_1 - \hat{p}_2$ tiene una distribución normal. La media y la desviación estándar del estadístico $\hat{p}_1 - \hat{p}_2$ son:

Media: $\hat{p}_1 - \hat{p}_2$

Desviación estándar: $\sqrt{(\hat{p}_1 \cdot \hat{q}_1/n_1) + (\hat{p}_2 \cdot \hat{q}_2/n_2)}$

Cuando se utilizan muestras grandes, la distribución muestral de la diferencia entre dos proporciones se puede calcular en forma aproximada a partir de la utilización de la distribución normal, mediante el estadístico Z, el cual se puede establecer a partir de:

$$Z = (\hat{p}_1 - \hat{p}_2) - (p_1 - p_2) / \sqrt{(\hat{p}_1 \cdot \hat{q}_1/n_1) + (\hat{p}_2 \cdot \hat{q}_2/n_2)}$$

El estadístico Z está distribuido en un intervalo que va de:

$$-Z_{\infty/2}, Z_{\infty/2} - Z_{\infty/2} < Z < Z_{\infty/2},$$

Sustituyendo el valor de Z se tiene:

$$-Z_{\infty/2} < (\hat{p}_1 - \hat{p}_2) - (p_1 - p_2) / \sqrt{(\hat{p}_1 \cdot \hat{q}_1/n_1) + (\hat{p}_2 \cdot \hat{q}_2/n_2)} < Z_{\infty/2}$$

Despejando $p_1 - p_2$ y resolviendo algebraicamente se obtiene:

$$(\hat{p}_1 - \hat{p}_2) - Z_{\infty/2} \cdot (\sqrt{\hat{p}_1 \cdot \hat{q}_1/n_1 + \hat{p}_2 \cdot \hat{q}_2/n_2}) < p_1 - p_2 < \hat{p}_1 - \hat{p}_2 + Z_{\infty/2} \cdot (\sqrt{\hat{p}_1 \cdot \hat{q}_1/n_1 + \hat{p}_2 \cdot \hat{q}_2/n_2})$$

Que es el intervalo de confianza para la diferencia entre dos proporciones poblacionales, es decir, la probabilidad que se tiene de que la diferencia de proporciones se encuentre en dicho intervalo.

La estimación por intervalos es un procedimiento de la estadística inferencial mediante el cual se realizan cálculos con los datos de una muestra cuyos resultados son dos valores numéricos que definen un rango, intervalo o conjunto numérico que servirá para estimar el parámetro poblacional.

La estimación por intervalos tiene varias ventajas; una es que no ofrece un valor único, sino un rango donde es muy posible o muy probable que el parámetro poblacional se encuentre incluido. De esta manera se supera la limitación de los estimadores puntuales de que su resultado único varía de muestra en muestra; es decir, con la estimación por intervalos tenemos más probabilidad de acertar al verdadero valor poblacional.

La principal ventaja de la estimación por intervalos es que su resultado ofrece un nivel de confianza que permite conocer en cuánto le podemos creer o tenerle confianza al resultado obtenido de la estimación. Por esta razón, la estimación por intervalos también es conocida como estimación por intervalos de confianza, pues su nivel de confianza señala qué tan posible o qué tan probable es que el parámetro poblacional se encuentre incluido dentro del rango definido.

UNIDAD VII

TÉCNICAS DE MUESTREO Y ESTIMACIÓN DE PARÁMETROS POBLACIONALES



www.mawil.us

La hipótesis es un planteamiento que sirve de fundamento para un análisis que pueda permitir la comprobación de su validez. La decima de hipótesis es la posibilidad de ser probado un planteamiento hipotético como pudiera ser que el consumo de un producto X favorece el crecimiento de los niños. Esta afirmación debe ser comprobada mediante una investigación que se desarrolle a tal efecto.

Los problemas de pruebas de hipótesis están muy relacionados con los estudios comparativos. Como se conoce en las hipótesis casi siempre se establecen comparaciones, que se refieren a valores o características poblacionales. El proceso de aceptación de una hipótesis en el terreno de la Estadística se hace a tenor de los valores que son observados en una determinada muestra o población.

Los tipos de hipótesis que establecen sus proposiciones son la hipótesis de nulidad y la hipótesis alternativa. Ambas se contraponen entre sí; la hipótesis de nulidad, es una hipótesis de diferencias nulas, que se establece de manera expresa para ser rechazada y que por consiguiente se acepte como verdadera la hipótesis alterna, que se convierte en la hipótesis de investigación.

$H_0: \theta = \theta_0$ (Hipótesis nula)

$H_1: \theta \neq \theta_0$ (Hipótesis alternativa)

Errores tipo I y II en la prueba de hipótesis

Cuando se debe tomar una decisión con respecto a la hipótesis de nulidad (H_0) existe la posibilidad de cometer dos tipos de errores. Se reconoce como error tipo I, el rechazo de H_0 siendo verdadera, el error tipo II, es lo contrario, aceptar H_0 siendo falsa.

La probabilidad de cometer el error tipo I está dada por α , en la medida en que sus valores aumenten es más probable rechazar a H_0 de manera equivocada. El error tipo II usualmente es representado por la letra β . Cuando las condiciones son ideales los valores y deberían estar precisados por el investigador. Esos valores determinan el tamaño de la muestra (N).

En la práctica es frecuente que α y N queden previamente definidos y luego se determina β . Entre α y β existe una relación inversamente proporcional.

Con el objetivo de disminuir las probabilidades de cometer los errores descritos se recurre a la función de potencia de una prueba estadística, la cual se define como la probabilidad de rechazar a H_0 cuando es verdaderamente falsa. Esto se representa:

$$\text{Potencia} = 1 - \text{probabilidad del error tipo II} = 1 - \beta.$$

Décima sobre la media. Casos. Cuando $n > 30$ y $n \leq 30$. Aplicaciones.

1. Cuando la varianza (s^2) es conocida ó $n \geq 30$

Se formula la siguiente Hipótesis :

$$H_0: \mu = \mu_0 \quad H_1: \mu \neq \mu_0 \quad \text{ó} \quad \mu < \mu_0 \quad \text{ó} \quad \mu > \mu_0$$

Función pibotal o experimental:

$$F_z = (X - \mu) / (\sigma / \sqrt{n})$$

Donde F_z sigue una distribución Normal.

2. Cuando la varianza (σ^2) es desconocida y $n \leq 30$

En este caso se formula la siguiente hipótesis:

$$H_0: \mu = \mu_0 \quad H_1: \mu \neq \mu_0 \quad \text{ó} \quad \mu < \mu_0 \quad \text{ó} \quad \mu > \mu_0$$

Función experimental

$$F_z = (X - \mu) / (s / \sqrt{n})$$

Donde t sigue una distribución t de Student con $n - 1$ grados de libertad.

Décima de hipótesis para la proporción poblacional.

1. Cuando se refiere a una población

La prueba de hipótesis para proporciones (porcentajes o probabilidad) es igual a una constante determinada.

Haciendo uso de la aproximación de la distribución normal a la binomial para muestras grandes se obtiene:

$$H_0: P = P_0$$

$$H_1: P \neq P_0, P < P_0, P > P_0$$

Obtenemos el estadístico

$$F_z = (X - n P_0) / (\sqrt{n P_0} (1 - P_0))$$

Donde: n tamaño de la muestra

2. Cuando se refiere a dos poblaciones (diferencia de proporciones). En este caso x_1, x_2 son variables aleatorias de dos poblaciones normales y se desea hacer una d6cima de hip6tesis para la igualdad de proporciones, se realiza mediante la funci3n experimental :

$$F_z = (p_1 - p_2) / (\sqrt{P(1-P)(1/n_1 + 1/n_2)}) : P = (X_1 + X_2) / (n_1 + n_2)$$

para la hip6tesis:

$$H_0: P_1 = P_2$$

$$H_1: P_1 \neq P_2, P_1 < P_2, P_1 > P_2$$

D6cima de hip6tesis para la varianza poblacional.

Si $x_1, x_2, x_3, \dots, x_n$ es una muestra aleatoria de una distribuci3n normal con $E[x] = m$ y $v[x] = \sigma^2$

$$H_0: \sigma^2 = \sigma_0^2$$

$$H_1: \sigma^2 \neq \sigma_0^2, \sigma^2 < \sigma_0^2, \sigma^2 > \sigma_0^2$$

La funci3n experimental est1 dado por:

$$F_{\chi^2} = ((n-1) S^2) / (\sigma_0^2)$$

Es el valor de una variable aleatoria χ^2 con $n - 1$ grados de libertad.

- 7.3 D6cima para la diferencia de medias para poblaciones independientes: Cuando $n > 30$ y cuando $n \leq 30$ si las varianza iguales y diferentes. Cuando las varianzas (σ_1^2) y (σ_2^2) son conocidas y cuando las varianzas (σ_1^2) y (σ_2^2) son desconocidas y $n_1, n_2 < 30$.

D6cima de hip6tesis de sobre la distribuci3n de una variable o Bondad de

Ajuste. Distribución Uniforme y Normal. Dócima de hipótesis respecto a las proporciones múltiples. Uso de software.

Dócima para la diferencia de medias para poblaciones independientes

1. Prueba de hipótesis para la diferencia de medias

1.1 Cuando las varianzas (σ_1^2) y (σ_2^2) son conocidas. En el caso que se quiera hacer comparaciones de los promedios de dos poblaciones:

$$H_0: \mu_1 = \mu_2$$

$$H_1: \mu_1 \neq \mu_2, \mu_1 < \mu_2, \mu_1 > \mu_2$$

De acuerdo a la alternativa se construye la región crítica.

La función experimental es:

$$F_z = ((X_1 - X_2) - (\mu_1 - \mu_2)) / \sqrt{((\sigma_1^2)/n_1 + (\sigma_2^2)/n_2)}$$

Donde F_z sigue una distribución normal estándar.

1.2) Cuando las varianzas (σ_1^2) y (σ_2^2) son desconocidas y $n_1, n_2 < 30$

Se usa la función experimental. \sim

$$F_z = ((X_1 - X_2) - (\mu_1 - \mu_2)) / (S_p \sqrt{(1/n_1 + 1/n_2)}) = ((X_1 - X_2) - (\mu_1 - \mu_2)) / \sqrt{((S_1^2)/n_1 + (S_2^2)/n_2)}$$

$$S_p = ((n_1 - 1)S_1^2 + (n_2 - 1)S_2^2) / (n_1 + n_2 - 2)$$

Es un t de student con $n_1 + n_2 - 2$ grados de libertad.

Dócima de hipótesis de sobre la distribución de una variable o Bondad de Ajuste. Distribución Uniforme y Normal.

La bondad de ajuste de un modelo estadístico representa lo adecuadamente que se ajusta un cúmulo de observaciones, por lo que se puede plantear que las medidas de bondad mayormente resumen la diferencia entre los valores observados y los valores esperados en el modelo de estudio, lo que permite su empleo en el contraste de hipótesis, como se puede observar en el test de normalidad de los residuos para comprobar si dos muestras se obtienen a partir de dos distribuciones idénticas o si las frecuencias siguen una distribución específica. En

el primer caso se puede emplear el estadístico de Kolmogorov-Smirnov y en el segundo caso se puede consultar la prueba de Chi-cuadrado.

- 7.4. Pruebas no paramétricas: Definición y características. Prueba U de Mann-Whitney. Prueba de H de Kruskal- Wallis. Prueba de Correlación por rangos de Spearman. Pruebas de ajuste de Kolmogorov-Smirnov.

Prueba U de Mann-Whitney

Las pruebas no-paramétricas se necesitan cuando no se tienen conocimientos con relación a la composición de los datos poblacionales, por lo tanto se desconoce sobre la distribución de probabilidad. Son empleadas cuando no se toman en cuenta las condiciones requeridas para la aplicación de las pruebas paramétricas. También son utilizadas en el caso de muestras pequeñas y carencia de información en relación a la densidad de probabilidad.

Dentro de las pruebas no-paramétricas, una muy utilizada es la Prueba de Suma de Rangos de Wilcoxon o también llamada prueba “U” de Mann-Whitney, la cual es una alternativa a la paramétrica “t” bimuestral. En estadística la prueba U de Mann-Whitneyes de tipo no paramétrica aplicada a dos muestras independientes. Es, en realidad, la versión no paramétrica de la conocida prueba “t” de Student. Frank Wilcoxon, la planteó en 1945 para muestras de igual tamaños y fue extendida a muestras de tamaño arbitrario como en otros sentidos por Henry B. Mann y D. R. Whitney en 1947.

Planteamiento de la prueba

La prueba de Mann-Whitney sirve para comprobar la heterogeneidad de dos muestras ordinales. El planteamiento inicial es que las observaciones de ambos grupos son independientes, son variables ordinales o continuas.

La hipótesis de nulidad es que la distribución de partida de ambos grupos es la misma, en tanto, la hipótesis alternativa, afirma que los valores de una de las muestras tienden a exceder a los de la otra:

$$P(X > Y) + 0.05 P(X = Y) > 0.05.$$

Cálculo estadístico

Para calcular el estadístico U se asigna a cada uno de los valores de las dos muestras su rango para construir

$$m_u = n_1 n_2 / 2$$
$$\sigma_u = (\sqrt{n_1 n_2 (n_1 + n_2 + 1) / 12})$$

Donde n_1 y n_2 son los tamaños respectivos de cada muestra; R_1 y R_2 es la suma de los rangos de las observaciones de las muestras 1 y 2 respectivamente.

El estadístico U se define como el mínimo de U_1 y U_2 .

Los cálculos tienen que tener en cuenta la presencia de observaciones idénticas a la hora de ordenarlas. No obstante, si su número es pequeño, se puede ignorar esa circunstancia.

Distribución del estadístico

La prueba calcula el llamado estadístico U, cuya distribución para muestras con más de 20 observaciones se aproxima bastante bien a la distribución normal.

La aproximación a la normal, z, cuando tenemos muestras lo suficientemente grandes viene dada por la expresión:

Donde m_U y s_U son la media y la desviación estándar de U si la hipótesis nula es cierta, y vienen dadas por las siguientes fórmulas:

Esta prueba estadística es ventajosa cuando las medidas se pueden disponer en escala ordinal (es decir, cuando los valores tienden a una variable continua, pero no tienen una distribución normal) y resulta aplicable cuando las muestras son independientes.

Este procedimiento es una buena opción, según se explicaba, cuando no hay posibilidad de aplicar la prueba t de Student, por no cumplir con los requisitos que esta prueba exige.

Prueba de H de Kruskal- Wallis

El análisis de varianza de una clasificación por rangos de Kruskal-Wallis es una prueba extremadamente útil para decidir si k muestras independientes son de poblaciones diferentes. Los valores de las muestras, casi invariablemente, difieren un poco y la cuestión radica en que la diferencias entre las muestras de signifiquen diferencia genuinas de población o simples variaciones aleatorias, semejantes a las esperadas entre distintas muestras aleatorias de la misma población. La técnica de Kruskal-Wallis examina la hipótesis de nulidad que suponen que las k muestras proceden de la misma población o de poblaciones idénticas con respecto a los promedios. La prueba supone que la variable en estudio tiene como base una distribución continua. Requiere, por lo menos, una medida ordinal de la variable.

Al calcular la prueba de Kruskal-Wallis, cada una de las N observaciones es reemplazada por rangos. Esto es, todos los puntajes de las k muestras combinadas se ordenan en una sola serie. El puntaje más pequeño es reemplazado por el rango 1, el siguiente en tamaño por el rango 2 y el más grande por el rango N. N es el número total de observaciones independientes en las k muestras.

Cuando se ha hecho esto, se encuentra la suma de los rangos de cada muestra (columna). La prueba de Kruskal-Wallis determina si la desigualdad entre las sumas de rangos es tan grande que probablemente no proceden de muestras tomadas de la misma población.

Puede demostrarse que si las k muestras efectivamente proceden de la misma población o de poblaciones idénticas, es decir, si H_0 es verdadera, H (estadística usada en la prueba de Kruskal-Wallis) está distribuida como chi-cuadrada con $df=k-1$, siempre que los tamaño de las diferentes k muestras no sean demasiado pequeñas. Así tenemos la fórmula:

$$H = 12/N(N + 1) \sum_{i=1}^k \frac{R_j^2}{n_j} - 3(N + 1)$$

k = número de muestras

n_j = número de casos en la muestra de orden j

$N = \sum n_j$, el número de casos de todas las muestras combinadas.

R_j = suma de rangos en la muestra de orden j

$\sum_{j=1}^k$ = indica sumar las k muestras (columnas)

Su distribución se aproxima a Chi-cuadrada con $df=k-1$, para tamaños muestrales (n_j) suficientemente grandes.

Prueba de Coeficiente de Correlación por rangos de Spearman (r_s)

De todas las estadísticas basadas en rangos, el coeficiente de Correlación de Spearman fue la primera que se desarrolló y es quizás el mejor conocido actualmente. Esta estadística, a veces llamada rho, se representa aquí con r_s . Es una medida de asociación que requiere que ambas variables sean medidas por lo menos en una escala ordinal, de manera que los objetos o individuos en estudio puedan colocarse en dos series ordenadas.¹⁰

Esto quiere decir que ambas variables X y Y se ordenan en una escala, donde la X representa las puntuaciones de los estudiantes en el examen de matemática y la Y los resultados de esos estudiantes en la prueba de español. De esta manera quedarán ordenados en $X_1, X_2, X_3, \dots, X_n$ (examen de matemática), y en $Y_1, Y_2, Y_3, \dots, Y_n$ (examen de español). Por tanto José representa X_1 en la prueba de matemática y Y_1 , en la prueba de español.

Para establecer la correlación entre ambas pruebas tenemos que buscar la diferencia (d) entre X y Y . En el caso de José sería:

$$d_1 = X_1 - Y_1$$

La correlación sería perfecta si y solo si $X_i=Y_i$ para todas las i . Las diferencias indicarán la disparidad entre ambos conjuntos.

Como el resultado de d_i puede variar en sus valores y obtener un signo negativo, este afectaría el cálculo del coeficiente de correlación por lo que todas las d_i son elevadas al cuadrado d_i^2 . Cuanto mayor sean las d_i , tanto menos perfecta será la asociación entre las variables.

La fórmula para calcular r_s es muy sencilla

Tenemos que $x = X - \bar{X}$, donde \bar{X} es la media de los puntajes de la variable X , entonces $y = Y - \bar{Y}$, de lo cual resulta como expresión general de un coeficiente de correlación la siguiente fórmula:

$$r_s = \frac{\sum xy}{\sqrt{(\sum x^2 \sum y^2)}}$$

Si la proporción de ligas en las observaciones de X o de Y es grande, se usa la fórmula:

$$r_s = \frac{(\sum x^2 + \sum y^2 - \sum d^2)}{2 \sqrt{(\sum x^2 \sum y^2)}}$$

En los restantes casos la fórmula más conveniente para calcular la r_s de Spearman es:

$$r_s = 1 - \frac{(\sum_{i=1}^N d_i^2)}{(N^3 - N)}$$

Prueba de ajuste de Kolmogorov-Smirnov.

Esta es una prueba de la bondad del ajuste o lo que es lo mismo, se interesa en el grado de acuerdo entre la distribución de un conjunto de valores de la muestra (puntajes observados) y alguna distribución teórica específica. Comprueba si racionalmente es probable que los puntajes obtenidos en la muestra deriven de una población que tenga esa distribución teórica.

Esta prueba contiene brevemente la especificación de la distribución de frecuencia acumulativa que ocurriría bajo la distribución teórica y su comparación

con la frecuencia acumulativa observada. La distribución teórica representa lo esperado conforme a H_0 . Se revela el punto en que ambas distribuciones presentan mayor divergencia. La distribución muestral indica que una divergencia de la magnitud observada ocurriría si las observaciones fueran realmente una muestra aleatoria de la distribución teórica.

El procedimiento a seguir es el siguiente:

$F_0(X)$ es la función de distribución de frecuencia acumulativa, cuyo valor es la proporción de casos esperados que tiene valores iguales o menores que X .

$S_n(X)$, es la distribución de frecuencia acumulativa observada de una muestra tomada al azar de N observaciones. En caso de que X es cualquier valor posible, tenemos $S_n(X) = k/N$, donde k es el número de observaciones igual o menor a X .

De acuerdo con la hipótesis de nulidad que supone una muestra obtenida de una distribución teórica específica, se espera que para cualquier valor de X , $S_n(X)$ se acerque claramente a $F_0(X)$. Conforme a la hipótesis de nulidad (H_0) se espera que la diferencia entre $S_n(X)$ y $F_0(X)$ sea pequeña y se encuentre en los límites de los errores aleatorios. La prueba Kolmogorov-Smirnov atiende la mayor de las desviaciones.

Al valor más grande de $F_0(X) - S_n(X)$ se le llama máxima desviación y se le representa con la letra D .

$$D = \text{máxima } [F_0(X) - S_n(X)]$$

UNIDAD VIII

ANALISIS DE VARIANZA (ANOVA)

DE K MEDIAS Y REGRESIÓN MÚLTIPLE



www.mawil.us

8.1 Análisis de varianza (ANOVA) Décima de hipótesis para la igualdad de K medias. El Análisis de la Varianza puede contemplarse como un caso especial de la modelización econométrica, donde el conjunto de variables explicativas son variables ficticias y la variable dependiente es de tipo continuo. En tales situaciones la estimación del modelo significa la realización de un análisis de la varianza clásica (ANOVA), de amplia tradición en los estudios y diseños experimentales.¹⁰

Estos análisis están relacionados con la necesidad que puede tener un investigador en encontrar similitudes en el comportamiento de un fenómeno que ocurre en dos o más situaciones, acontecimientos, grupos, etc.

Una manera sencilla estaría en la comparación de las medias de estos grupos y ver si las medias aritméticas de la variable estudiada son parecidas o diferentes, sin embargo esto no reúne la validez necesaria por la influencia que ejerce la dispersión de las observaciones.

En los análisis de varianza es posible encontrar con múltiples factores a estudiar paralelamente. Del mismo modo, podemos diferenciar tres tipos de modelos según sean de:

- Efectos fijos: donde sólo estudiamos determinados niveles del factor (ejemplo el tamaño: alto, mediano y bajo) y únicamente perseguimos sacar conclusiones para éstos.
- Efectos aleatorios: en este caso los niveles son infinitos y estudiamos una muestra de los mismos. Sus resultados también serán aleatorios.
- Efectos mixtos: están incluidos uno o más factores de las clases anteriores.

ANOVA es particularmente útil cuando deseamos analizar distintas situaciones o alternativas de actuación en las que podemos intervenir en la realización del experimento. A diferencia del análisis econométrico habitual, donde las series históricas son dadas y no podemos repetir la situación, ni modificar alguna de las condiciones o variables (pensemos en el P.I.B., inflación, etc.) para estudiar sus efectos, en el contexto ANOVA nos encontraremos la mayoría de las

veces ante datos experimentales (controlables y/o repetibles en mayor o menor grado).

El modelo ANOVA tradicional tiene la expresión:

$$Y_{ij} = \mu + \tau_j + \varepsilon_{ij}$$

Y es la variable objeto de estudio y que en nuestro caso es la venta para el establecimiento i del nivel j.

μ = es una constante e indica la respuesta media de todos los niveles.

τ_j = es el efecto diferencial del nivel j. Recoge la importancia de cada tratamiento y es el objetivo del análisis. Dado que los τ_j son efectos diferenciales sobre μ tenemos que $\sum \tau_j = 0$

ε_{ij} = es un término de error, considerado como variable aleatoria $N \sim (0, \sigma_2)$

El ANOVA tradicional parte de descomponer la variación total de la muestra, en dos componentes:

VARIACIÓN TOTAL = VARIACIÓN ENTRE + VARIACIÓN INTRA

Esta igualdad básica nos indica que la variación total es igual a la suma de la variación o dispersión entre los grupos, más la variación o dispersión dentro de cada grupo. Los grupos están definidos por los niveles de factor.

La anterior igualdad puede expresarse por:

$$\underbrace{\sum_{j=1}^g \sum_{i=1}^{n_j} (Y_{ij} - \bar{Y}_{..})^2}_{V. \text{ TOTAL}} = \underbrace{\sum_{j=1}^g n_j (\bar{Y}_{.j} - \bar{Y}_{..})^2 + \sum_{j=1}^g}_{V. \text{ ENTRE}} \underbrace{\sum_{i=1}^{n_j} (Y_{ij} - \bar{Y}_{.j})^2}_{V. \text{ INTRA}}$$

Correspondiendo cada término de la suma a las anteriores variaciones y siendo la $\bar{Y}_{..}$ media total e $\bar{Y}_{.j}$ la media de grupo o nivel j.

Los grados de libertad (número de observaciones –parámetros a estimar) correspondientes a cada uno de los componentes de la variación total son:

- Variación ENTRE: $g - 1$
- Variación INTRA: $n - g$
- Variación TOTAL: $n - 1$

Dado que a través del Análisis de la Varianza se persigue saber si los distintos niveles de un factor influye en los valores de una variable continua para que efectivamente sí haya diferencias en los valores de la variable continua según el nivel del factor, se tiene que dar simultáneamente que el comportamiento de la variable continua sea lo más distinto posible para los distintos niveles del factor, y a su vez, que dentro de cada grupo (determinado por los niveles del factor) los valores sean lo más homogéneos posibles.

En otras palabras, se tiene que dar que la variación intragrupos sea mínima, y que la variación entre grupos sea máxima.

Por tanto, el análisis de la varianza se va a basar no sólo en la descomposición de la variación total, sino además en la comparación de la variación ENTRE - grupos y la variación INTRA- grupos, teniendo en cuenta sus correspondientes grados de libertad.

Introducción Regresión Simple. Regresión Múltiple. Casos de k variables independientes. Estimación de los coeficientes de regresión.

Un análisis de regresión genera una ecuación para describir la relación estadística entre uno o más predictores y la variable de respuesta y para predecir nuevas observaciones. La regresión lineal generalmente utiliza el método de estimación de mínimos cuadrados ordinarios, del cual se obtiene la ecuación al minimizar la suma de los residuos al cuadrado.

Los resultados de la regresión identifican la dirección, el tamaño y la significancia estadística de la relación entre un predictor y una respuesta.

- El signo de cada coeficiente indica la dirección de la relación.
- Los coeficientes representan el cambio medio en la respuesta para una unidad de cambio en el predictor mientras se mantienen constantes otros predictores incluidos en el modelo.
- El valor p de cada coeficiente prueba la hipótesis nula de que el coeficiente es igual a cero (sin efecto). Por lo tanto, los valores p bajos indican que el predictor es una adición significativa al modelo.
- La ecuación predice nuevas observaciones dados los valores predictores especificados.

Los modelos con un predictor se denominan regresión simple. Los modelos con más de un predictor se conocen como regresión lineal múltiple.

¿Qué es regresión lineal simple?

La regresión lineal simple analiza la relación lineal entre dos variables continuas: una respuesta (Y) y un predictor (X). Cuando las dos variables están vinculadas, es posible predecir un valor de respuesta a partir de un valor predictor con una exactitud mayor que la asociada únicamente a las probabilidades.

La regresión proporciona la línea que mejor se ajusta a los datos. Esta línea se puede utilizar después para:

- Examinar cómo cambia la variable de respuesta a medida que cambia la variable predictora.
- Predecir el valor de una variable de respuesta (Y) para cualquier variable predictora (X).

¿Qué es regresión lineal múltiple?

La regresión lineal múltiple examina las relaciones lineales entre una respuesta continua y dos o más predictores.

Si el número de predictores es grande, antes de ajustar un modelo de regresión con todos los predictores, se deberían utilizar las técnicas de selección de

modelo paso a paso o de los mejores subconjuntos para excluir los predictores que no estén asociados con las respuestas.

¿Qué es la regresión de mínimos cuadrados ordinarios?

En la regresión de mínimos cuadrados ordinarios (OLS), la ecuación estimada se calcula determinando la ecuación que minimiza la suma de las distancias al cuadrado entre los puntos de los datos de la muestra y los valores pronosticados por la ecuación.

Con un predictor (regresión lineal simple), la suma de las distancias al cuadrado desde cada punto hasta la línea deben ser tan pequeñas como sea posible.

Supuestos deben cumplirse para regresión OLS.

La regresión OLS proporciona las estimaciones sin sesgo más precisas solo cuando se cumplen los siguientes supuestos:

1. El modelo de regresión es lineal en los coeficientes.
 2. Los mínimos cuadrados pueden modelar la curvatura al transformar las variables (en lugar de los coeficientes).
 3. Se debe especificar la forma funcional correcta para poder modelar cualquier curvatura. En este caso, la variable predictora, X , se eleva al cuadrado para modelar la curvatura. $Y = b_0 + b_1X + b_2X^2$
- Los residuos tienen una media de cero. La inclusión de una constante en el modelo hará que la media sea igual a cero.
 - Todos los predictores no están correlacionados con los residuos.
 - Los residuos no están correlacionados entre sí (correlación en serie).
 - Los residuos tienen una varianza constante.
 - Ninguna variable predictora está correlacionada perfectamente ($r=1$) con una variable predictora diferente. También es mejor evitar las correlaciones imperfectamente altas (multicolinealidad).
 - Los residuos están distribuidos normalmente.

Puesto que la regresión OLS proporcionará las mejores estimaciones solo cuando se cumplan todos estos supuestos, es muy importante evaluarlos. Los métodos comunes incluyen examinar gráficas de residuos, usar pruebas de falta de ajuste y ver la correlación entre los predictores usando el factor de inflación de la varianza (FIV).

BIBLIOGRAFÍA

ANÁLISIS DE VARIANZA (ANOVA)

DE K MEDIAS Y REGRESIÓN MÚLTIPLE.



www.mawil.us

1. Fuente, S.(2016) Estadística teórica: distribuciones de probabilidad. Universidad Autónoma de Madrid. Facultad Ciencias Económicas y Empresariales. Disponible en:<http://www.fuenterrebollo.com/Aeronautica2016/Estadisticateórica.distribuciones.pdf>
2. Fuente, S.(2016) Estimadores.Propiedades.Aplicación. Universidad Autónoma de Madrid. Facultad Ciencias Económicas y Empresariales.Disponible en:<http://www.estadistica.net/Aeronautica2016/estimadores.pdf>
3. Ramirez, O.(2012).Media y varianza de una distribución discreta de probabilidad. Disponible en:<http://roa.uveg.edu.mx/repositorio/licenciatura/47/Lectura2Mediayvarianzadeuadistribucindiscretadeprobabilidad.pdf>
4. Vicens, J; Herrarte, A; Medina,E.(2005) ANÁLISIS DE LA VARIANZA(ANOVA)Disponible en: <http://www.uam.es/departamentos/economicas/econapli/anova.pdf>
5. Cavada, G.(2009). Docimasia de hipótesis.Rincón de la Bioestadística. Rev. chil. endocrinol.diabetes 2009; 2 (4): 256-257Disponible en:http://www.revistasoched.cl/4_2009/13-4-2009.pdf
6. Valdivieso, C; Valdivieso, R; Valdivieso, O.(2011) DETERMINACIÓN DEL TAMAÑO MUESTRAL MEDIANTE EL USO DE ÁRBOLES DE DECISIÓN. UPB- Investigación y desarrolloDisponible en:<ftp://ftp-repec.org/opt/ReDIF/RePEc/iad/wpaper/0311.pdf>
7. Dicovskiy, L.(2008).Estadística Básica. UNIVERSIDAD NACIONAL DE INGENIERÍA. Nicaragua. Disponible en: http://frrq.cvg.utn.edu.ar/pluginfile.php/2101/mod_resource/content/0/DEPOSITO_DE_MATERIALES/estadistica1_1_.pdf
8. Rohen, V. ESTIMACIÓN. Puntual y por intervalo. Disponible en: <http://lcolladotor.github.io/courses/Courses/MEyAddDG/day2/Estimaci%C3%B3n.pdf>
9. Faraldo,P; Pateiro,B(2013)Metodología de la Investigación.Estimación de parámetros.Universidad de Santiago de Compostela. Documento pdf.
- 10.González, B; Hernández,D; Jiménez,M; Marrero,I; Sanabria,A(2013). Muestreo y estimación. Matemática Aplicada y Estadística. Disponible en: https://campusvirtual.ull.es/ocw/pluginfile.php/6115/mod_resource/content/1/tema/ME9-muestreo.pdf
- 11.Siegel,S (1982). Estadística no paramétrica aplicada a las ciencias de la conducta. México, D.F.Editorial Trillas. Séptima reimpresión.

12. Wiersma, W. (1986) Research methods in education: an introduction. Boston, Mass: Allyn and Bacon.

ESTADÍSTICAS PARA NEGOCIOS INTERNACIONALES



Publicado en Ecuador
Marzo del 2019

Edición realizada desde el mes de agosto del año 2018 hasta octubre del año 2018, en los talleres Editoriales de MAWIL publicaciones impresas y digitales de la ciudad de Quito.

Quito – Ecuador

Tiraje Digital



Estadísticas para Negocios Internacionales



EC. LUIS AGUSTIN FAJARDO VACA MG.
MAGISTER EN ADMINISTRACIÓN DE PEQUEÑAS Y MEDIANAS EMPRESAS;
ECONOMISTA; LICENCIADO EN CIENCIAS POLÍTICAS Y SOCIALES
DOCENTE UNIVERSIDAD DE GUAYAQUIL
LUIS.FAJARDOVA@UG.EDU.EC



ING. JOHANNA ESTEFANIA RANGEL SALTOS MG.
MAGISTER EN CIENCIAS CON MENCIÓN EN ECONOMÍA Y GESTIÓN EMPRESARIAL;
INGENIERA EN GESTIÓN EMPRESARIAL INTERNACIONAL
DOCENTE UNIVERSIDAD DE GUAYAQUIL
JOHANNA.RANGELS@UG.EDU.EC



EC. NANCY ROCIO CASTILLO CASTRO MG.
MAGISTER EN FINANZAS Y PROYECTOS CORPORATIVOS; ECONOMISTA
DOCENTE UNIVERSIDAD DE GUAYAQUIL
NANCY.CASTILLOC@UG.EDU.EC



ING. KAREN GISELLE OCHOA QUIROLA MG.
MAGISTER EN ADMINISTRACIÓN DE EMPRESAS CON MENCIÓN EN
RECURSOS HUMANOS; INGENIERA EN GESTIÓN EMPRESARIAL
DOCENTE UNIVERSIDAD DE GUAYAQUIL
KAREN.OCHOAQ@UG.EDU.EC

ISBN: 978-9942-787-57-6



9 789942 787576

