BIOESTADISTICA

Septiembre – Diciembre

Marco Estratégico de Referencia

Antecedentes históricos

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

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

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

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

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

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

Misión

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

Visión

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

Valores

- Disciplina
- Honestidad
- Equidad
- Libertad

Escudo

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

Eslogan

“Pasión por Educar”

Balam

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

BIOESTADISTICA

Objetivo de la materia:

Al finalizar la asignatura el alumno conocerá las definiciones básicas de la estadística descriptiva y su aplicación en el campo de la enfermería, manejará adecuadamente las técnicas de estimación y decisión en la realización de estudios de corte epidemiológico y en las investigaciones propias de la disciplina.

UNIDAD I

ESTADÍSTICA DESCRIPTIVA

1.1 La estadística en enfermería.

1.1.1 Introducción histórica.

1.1.2 Elementos del análisis estadístico en enfermería.

1.1.3. La estadística como herramienta de trabajo en enfermería.

1.2. Descripción de una variable estadística.

1.2.1. Definiciones básicas.

1.2.2. Representaciones gráficas.

1.2.3. Representación numérica.

1.2.4. Características de posición, dispersión y forma.

1.3. Descripción numérica de una variable estadística bidimensional.

1.3.1 Distribuciones marginales y condicionadas.

1.3.2. Independencia e incorrelación.

1.3.3. Características numéricas.

1.4. Regresión y correlación.

1.4.1. Definiciones.

1.4.2. Curva de regresión y coeficiente de determinación.

3.1.3. Estadísticos muéstrales. Distribuciones.

3.2. Estimación.

3.2.1. Propiedades de los estimadores.

3.2.2. Obtención de estimadores.

3.2.3. Estimación por intervalos de confianza.

3.3. Contraste de hipótesis.

3.3.1. Concepto y definiciones.

3.3.2. Construcción de Test de hipótesis.

3.4. Contraste de hipótesis paramétricas.

3.4.1. Test para poblaciones normales.

3.4.2. Test para poblaciones binomiales y de Poisson.

3.5. Test basado en el estadístico chi-cuadrado.

3.5.1. Test de bondad de ajuste.

3.5.2. Test de heterogeneidad.

3.5.3. Test de homogeneidad.

3.5.4. Tablas de Contingencia.

UNIDAD IV

4.1. Conceptos de demografía.

4.1.1. Conceptos básicos.

4.1.2. Modelos de crecimiento de poblaciones.

4.1.3. Fuentes Históricas. Fuentes Naturales.

4.1.4 Fenómenos Demográficos.

Criterios de evaluación:

No	Concepto	Porcentaje
1	Trabajos Escritos	30%
2	Actividades áulicas	20%
3	Examen	50%
4	Total	100%
5	Escala de calificación	7- 10
6	Mínima aprobatoria	7

UNIDAD I

ESTADÍSTICA DESCRIPTIVA

1.1 La estadística en enfermería.

Aunque aparentemente la bioestadística parece una ciencia fundamentalmente teórica, es utilizada en la práctica médica a diario. Cuando hablamos de la dosis media de eritropoyetina administrada en diálisis o el tiempo medio de duración de una sesión de hemodiálisis estamos utilizando la estadística. O cuando decidimos tratar a un paciente con unas cifras de colesterol o de presión arterial elevadas, previamente se ha demostrado estadísticamente que existe un riesgo elevado cuando esas cifras están por encima de un determinado valor. O por ejemplo, cuando queremos comparar si existe diferencias entre un fármaco inmunosupresor u otro a la hora de prevenir el rechazo de un trasplante renal.

El objetivo es iniciar y familiarizar a la enfermería con el método científico, y más concretamente con las nociones básicas del análisis estadístico necesario para cualquier estudio de investigación.

La bioestadística es una rama de la estadística que se ocupa de los problemas planteados dentro de las ciencias de la vida, como la biología, la medicina, la enfermería, entre otras.

La primera razón es que la información numérica está en todas partes. Por ejemplo, en los periódicos, revistas de noticias o de interés general, revistas de enfermería y de ciencias de la salud en general, informes de investigación en salud, noticias de televisión, radio, etc., se encuentra gran información numérica. Para ser consumidores educados en esta información, es necesario poder leer las tablas y gráficas, así como entender el análisis de la información numérica.

Una segunda razón es que las técnicas estadísticas se utilizan para tomar decisiones que afectan nuestra vida y nuestro ejercicio profesional.

Una tercera razón es que el conocimiento de los métodos estadísticos ayuda a entender cómo se toman las decisiones y a comprender de qué manera nos afectan a nivel personal, profesional, institucional y social.

En cualquier línea del trabajo enfermero es preciso tomar decisiones en las que el entendimiento del análisis de datos es de mucha utilidad.

La estadística nos va a ayudar a seleccionar las conclusiones generales más adecuadas a partir de datos parciales y representativos.

1.1.1 Introducción histórica.

El primer médico que utilizó métodos matemáticos para cuantificar variables de pacientes y sus enfermedades fue el francés Pierre Charles-Alexandre Louis (1787-1872). La primera aplicación del Método numérico (que es como tituló a su obra y llamó a su método) en su clásico estudio de la tuberculosis, que influyó en toda una generación de estudiantes. Sus discípulos, a su vez, reforzaron la nueva ciencia de la epidemiología con en el método estadístico. En las recomendaciones de Louis para evaluar diferentes métodos de tratamiento están las bases de los ensayos clínicos que se hicieron un siglo después. En Francia Louis René Villermé (1782-1863) y en Inglaterra William Farr (1807-1883) —que había estudiado estadística médica con Louis— hicieron los primeros mapas epidemiológicos usando métodos cuantitativos y análisis epidemiológicos. Francis Galton (1822-1911), basado en el darwinismo social, fundó la biometría estadística.

Pierre Simon Laplace (1749-1827), astrónomo y matemático francés, publicó en 1812 un tratado sobre la teoría analítica de las probabilidades, Théorie analytique des probabilités, sugiriendo que tal análisis podría ser una herramienta valiosa para resolver problemas médicos.

Los primeros intentos de hacer coincidir las matemáticas de la teoría estadística con los conceptos emergentes de la infección bacteriana tuvieron lugar a comienzos del siglo XX. Tres diferentes problemas cuantitativos fueron estudiados por otros tantos autores. William Heaton Hamer (1862-1936) propuso un modelo temporal discreto en un intento de explicar la ocurrencia regular de las epidemias de sarampión; John Brownlee (1868-1927), primer director del British Research Council, luchó durante veinte años con problemas de cuantificación de la infectividad epidemiológica, y Ronald Ross (1857-1932) exploró la aplicación matemática de la teoría de las probabilidades con la finalidad de determinar la relación entre el número de mosquitos y la incidencia de malaria en situaciones endémicas y epidémicas. Pero el cambio más radical en la dirección de la epidemiología se debe a Austin Bradford Hill (1897-1991) con el ensayo clínico aleatorizado y, en colaboración con Richard Doll (n. 1912), el épico trabajo que correlacionó el tabaco y el cáncer de pulmón.

Los primeros trabajos bioestadísticos en enfermería los realizó, a mediados del siglo XIX la enfermera inglesa Florence Nightingale. Durante la guerra de Crimea, Florence Nightingale observó que eran mucho más numerosas las bajas producidas en el hospital que en el frente. Por lo tanto, recopiló información y dedujo que la causa de la elevada tasa de mortalidad se debía a la precariedad higiénica existente. Así, gracias a sus análisis estadísticos, se comenzó a tomar conciencia de la importancia y la necesidad de unas buenas condiciones higiénicas en los hospitales.

1.1.2 Elementos del análisis estadístico en enfermería.

La estadística descriptiva comprende la presentación, organización y resumen de los datos de una manera científica. Incluye diversos métodos de organizar y representar gráficamente los datos, para dar una idea de lo que nos muestran. Las tablas, los diagramas de barras o los gráficos sectoriales o "tartas" son algunos de los elementos de estadística descriptiva. También incluye varios parámetros numéricos (como la media aritmética) que resumen los datos con muy pocos números clave. Por otra parte, la estadística inferencial o inductiva permite generalizar los datos obtenidos a partir de una muestra a un número mayor de individuos (población). La estadística inferencial se basa en la teoría de las probabilidades y trabaja con los datos que le proporciona la estadística descriptiva.

1.1.3. La estadística como herramienta de trabajo en enfermería.

En el mundo actual, Holmes (1980) señala que la Estadística es necesaria para que un ciudadano con educación general adquiera la capacidad de lectura e interpretación de tablas y gráficos estadísticos que aparecen en los medios informativos, de orientarse en un mundo ligado por las telecomunicaciones e interdependiente y de interpretar una amplia gama de información sobre los temas más variados. El conocimiento de la estadística favorece el desarrollo personal pues fomenta un razonamiento crítico, aumenta la capacidad de usar datos cuantitativos para controlar nuestros juicios e interpretar los ajenos y trasformarlos para resolver problemas de decisión y efectuar predicciones. (Ottaviani, 1998).En Enfermería el estudio de la Estadística aporta los conceptos fundamentales y necesarios con el dominio adecuado del instrumental para aproximarse al estudio y conocimiento de los fenómenos de competencia de la Enfermería. La práctica de la investigación y la transferencia de conocimientos producidos al ejercicio profesional, constituye la actividad básica para el desarrollo de la Enfermería a través del cual se aspira a la meta social de dar respuesta a los problemas y necesidades de la comunidad.En el campo de la Salud, las prioridades de investigación exigen que el personal que se forma y trabaja en el sector incorpore la investigación como una actividad permanente en su ámbito de acción. La Estadística desempeña un papel importante en la toma de decisiones en todas las áreas, entre ellas la salud pública. Las medidas relativas a diferentes programas sanitarios confían en parte, en las predicciones sobre la longevidad de la población, o cómo invertir recursos para reducir la mortalidad infantil, disminuir la probabilidad de muerte en accidentes vehiculares con eluso del cinturón de seguridad, cuáles factores incrementan el riesgo de que un individuo desarrolle una enfermedad coronaria.

1.2. Descripción de una variable estadística.

Una variable estadística es una característica que puede fluctuar y cuya variación es susceptible de adoptar diferentes valores, los cuales pueden medirse u observarse. Las variables adquieren valor cuando se relacionan con otras variables, es decir, si forman parte de una hipótesis o de una teoría. En este caso se las denomina constructos o construcciones hipotéticas.

1.2.1. Definiciones básicas.

Según el nivel de medición o también según el criterio metodológico,1 pueden ser:

Variables cualitativas

Son el tipo de variables que como su nombre lo indica expresan distintas cualidades, características o modalidad. Cada modalidad que se presenta se denomina atributo o categoría, y la medición consiste en una clasificación de dichos atributos. Las variables cualitativas pueden ser dicotómicas cuando sólo pueden tomar dos valores posibles, como sí y no, hombre y mujer o ser politómicas cuando pueden adquirir tres o más valores. Dentro de ellas podemos distinguir:

- Variable cualitativa ordinal o variable cuasicuantitativa: La variable puede tomar distintos valores ordenados siguiendo una escala establecida, aunque no es necesario que el intervalo entre mediciones sea uniforme, por ejemplo: leve, moderado, fuerte.

- Variable cualitativa nominal: En esta variable los valores no pueden ser sometidos a un criterio de orden, como por ejemplo los colores o el lugar de registro.

Variables cuantitativas

Son las variables que toman como argumento cantidades numéricas, son variables matemáticas. Las variables cuantitativas además pueden ser:

- Variable discreta: Es la variable que presenta separaciones o interrupciones en la escala de valores que puede tomar. Estas separaciones o interrupciones indican la ausencia de valores entre los distintos valores específicos que la variable pueda asumir. Ejemplo: El número de hijos (1, 2, 3, 4, 5).
- Variable continua: Es la variable que puede adquirir cualquier valor dentro de un intervalo especificado de valores. Por ejemplo la masa (2,3 kg, 2,4 kg, 2,5 kg,...) o la altura (1,64 m, 1,65 m, 1,66 m,...), o el salario. Solamente se está limitado por la precisión del aparato medidor, en teoría permiten que exista indefinidos valores entre dos variables.

Según la influencia que le asignemos a unas variables sobre otras, estas podrán ser:

Variables independientes

Una variable independiente es aquella cuyo valor no depende de otra variable. Es aquella característica o propiedad que se supone es la causa del fenómeno estudiado. En investigación experimental se llama así a la variable que el investigador manipula.

Las variables independientes son las que el investigador escoge para establecer agrupaciones en el estudio, clasificando intrínsecamente a los casos del mismo. Un tipo especial son las variables de control, que modifican al resto de las variables independientes y que de no tenerse en cuenta adecuadamente pueden alterar los resultados por medio de un sesgo.

La variable independiente se suele representar en el eje de abscisas.

La variable independiente es la que se le asignan valores arbitrarios

Variables dependientes

Una variable dependiente es aquella cuyos valores dependen de los que tomen otra variable. La variable dependiente es una función que se suele representar por la y. La variable dependiente se representa en el eje ordenadas. Son las variables de respuesta que se observan en el estudio, y que podrían estar influidas por los valores de las variables independientes.

La variable dependiente es el factor que es observado y medido para determinar el efecto de la variable independiente.

1.2.2. Representaciones gráficas.

Una gráfica o representación gráfica es un tipo de representación de datos, generalmente numéricos, mediante recursos visuales (líneas, vectores, superficies o símbolos), para que se manifieste visualmente la relación matemática o correlación estadística que guardan entre sí. También es el nombre de un conjunto de puntos que se plasman en coordenadas cartesianas y sirven para analizar el comportamiento de un proceso o un conjunto de elementos o signos que permiten la interpretación de un fenómeno. La representación gráfica permite establecer valores que no se han obtenido experimentalmente sino mediante la interpolación (lectura entre puntos) y la extrapolación (valores fuera del intervalo experimental).

1.2.3. Representación numérica.

La presentación de datos estadísticos constituye en sus diferentes modalidades uno de los aspectos de más uso en la estadística descriptiva. A partir podemos visualizar a través de los diferentes medios escritos y televisivos de comunicación masiva la presentación de los datos estadísticos sobre el comportamiento de las principales variables económicas y sociales, nacionales e internacionales.

1-Presentación escrita: Esta forma de presentación de informaciones se usa cuando una serie de datos incluye pocos valores, por lo cual resulta más apropiada la palabra escrita como forma de escribir el comportamiento de los datos; mediante la forma escrita, se resalta la importancia de las informaciones principales.

2-Presentación tabular: Cuando los datos estadísticos se presentan a través de un conjunto de filas y de columnas que responden a un ordenamiento lógico; es de gran eso e importancia para el uso e importancia para el usuario ya que constituye la forma más exacta de presentar las informaciones. Una tabla consta de varias partes, las principales son las siguientes:

Título: Es la parte más importante del cuadro y sirve para describir todo él contenido de este.

Encabezados: Son los diferentes subtítulos que se colocan en la parte superior de cada columna.

Columna matriz: Es la columna principal del cuadro.

Cuerpo: El cuerpo contiene todas las informaciones numéricas que aparecen en la tabla.

Fuente: La fuente de los datos contenidos en la tabla indica la procedencia de estos.

Notas al pie: Son usadas para hacer algunas aclaraciones sobre aspectos que aparecen en la tabla o cuadro y que no han sido explicados en otras partes.

1.2.4. Características de posición, dispersión y forma.

Las medidas de posición proporcionan información resumida de la variable objeto de estudio.

Medidas de posición centrales

- Media (aritmética, geométrica y armónica)
- Mediana
- Moda

Medidas de posición no centrales

- Cuantiles (cuartiles, deciles y percentiles).

Las medidas de dispersión estudian la separación existente entre los valores que toma la variable.

Medidas de dispersión absolutas

- Rango
- Recorrido intercuartílico
- Desviación absoluta media respecto a la media
- Varianza
- Desviación típica

Medidas de dispersión relativas

- Coeficiente de apertura
- Recorrido relativo
- Recorrido semi-intercuartílico
- Coeficiente de variación
- Variable tipificada

medidas de forma permiten comprobar si una distribución de frecuencia tiene características especiales como simetría, asimetría, nivel de concentración de datos y nivel de apuntamiento que la clasifiquen en un tipo particular de distribución.

Las medidas de forma son necesarias para determinar el comportamiento de los datos y así, poder adaptar herramientas para el análisis probabilístico.

1.3. Descripción numérica de una variable estadística bidimensional.

En numerosas ocasiones interesa estudiar simultáneamente dos (o más) caracteres de una población. En el caso de dos (o más) variables estudiadas conjuntamente se habla de variable bidimensional (multidimensional); si se trata de dos caracteres cualitativos, de par de atributos.

Si de una cierta población se estudian dos caracteres simultáneamente se obtienen dos series de datos.

Variable estadística bidimensional es el conjunto de pares de valores de dos caracteres o variables estadísticas unidimensionales X e Y sobre una misma población.

La variable estadística bidimensional se representa por el símbolo (X, Y) y cada uno de los individuos de la población viene caracterizado por la pareja (x_i , y_i), en el cual x_i representa los datos, valores o marcas de clase $x_1, x_2, ..., x_n$ de la variable X; e y_i representa los datos, valores o marcas de clase $y_1, y_2, ..., y_m$ de la variable Y.

Se denominan distribuciones bidimensionales a las tablas estadísticas bidimensionales formadas por todas las frecuencias absolutas de todos los posibles valores de la variable estadística bidimensional (X, Y).

Las tablas estadísticas bidimensionales pueden ser:

a) Simples.

b) De doble entrada.

1.3.1 Distribuciones marginales y condicionadas.

En teoría de probabilidades, la distribución marginal es la distribución de probabilidad de un subconjunto de variables aleatorias de un conjunto de variables aleatorias. La distribución marginal proporciona la probabilidad de un subconjunto de valores del conjunto sin necesidad de conocer los valores de las otras variables. Esto contrasta con la distribución condicional, que proporciona probabilidades contingentes sobre el valor conocido de otras variables.

El término variable marginal se usa para referirse a una variable del subconjunto de retenido y cuyos valores pueden ser conocidos.1 La distribución de las variables marginales, la distribución marginal, se obtiene marginalizando sobre la distribución de variables descartadas y las variables descartadas se llaman a veces variables marginalizadas.

Partiendo de una distribución bidimensional de frencuencias (xi, yj; nij), se denomina distribución condicionada de la variable X a un valor dado yj de la variable Y a la distribución unidimensional definida por el conjunto de valores tomados por X y de las frencuencias condicionadas de dichos valores de X a qué Y tome el valor yj.

Análogamente, se denomina distribución de la variable y condicionada a un valor dado xi de la variable X a la distribución unidimensional definida por el conjunto de valores tomados por Y y de las frecuencias de dichos valores de Y condicionadas a que X tome el valor xi.

1.3.2. Independencia e incorrelación.

Dos variables estadísticas son estadísticamente independientes cuando el comportamiento estadístico de una de ellas no se ve afectado por los valores que toma la otra; esto es cuando las relativas de las distribuciones condicionadas no se ven afectadas por la condición, y coinciden en todos los casos con las frecuencias relativas marginales.

Esta definición puede hacerse más operativa, a través de la caracterización siguiente: Dos variables son estadísticamente independientes cuando para todos los pares de valores se cumple que la frecuencia relativa conjunta es igual al producto de las frecuencias relativas marginales.

1.3.3. Características numéricas.

Los sistemas de numeración son conjuntos de dígitos usados para representar cantidades, así se tienen los sistemas de numeración decimal, binario, octal, hexadecimal, romano, etc. Los cuatro primeros se caracterizan por tener una base (número de dígitos diferentes: diez, dos, ocho, dieciseis respectivamente) mientras que el sistema romano no posee base y resulta más complicado su manejo tanto con números, así como en las operaciones básicas.

Los sistemas de numeración que poseen una base tienen la característica de cumplir con la notación posicional, es decir, la posición de cada número le da un valor o peso, así el primer dígito de derecha a izquierda después del punto decimal, tiene un valor igual a b veces el valor del dígito, y así el dígito tiene en la posición n un valor igual a: (bn) * A

donde:

b = valor de la base del sistema

n = número del dígito o posición del mismo

A = dígito.

1.4. Regresión y correlación.

En análisis de regresión es una herramienta de frecuente uso en Estadística que permite estudiar y valorar las relaciones entre diferentes variables cuantitativas tenidas en cuenta mediante la construcción de una ecuación.

El esquema básico de análisis de regresión plantea un proceso o modelo en el cual se analiza la relación entre una variable dependiente (porque es influida por el resto) y una o varias variables independientes o fijas (las que influyen en el objeto de estudio). Gracias procesos de regresión estadística es posible entender el modo en que la dependiente es afectada por cambios en los valores de las independientes.

La correlación estadística determina la relación o dependencia que existe entre las dos variables que intervienen en una distribución bidimensional.

Es decir, determinar si los cambios en una de las variables influyen en los cambios de la otra. En caso de que suceda, diremos que las variables están correlacionadas o que hay correlación entre ellas.

En probabilidad y estadística, la correlación indica la fuerza y la dirección de una relación lineal y proporcionalidad entre dos variables estadísticas. Se considera que dos variables cuantitativas están correlacionadas cuando los valores de una de ellas varían sistemáticamente con respecto a los valores homónimos de la otra: si tenemos dos variables (A y B) existe correlación entre ellas si al disminuir los valores de A lo hacen también los de B y viceversa. La correlación entre dos variables no implica, por sí misma, ninguna relación de causalidad.

1.4.1. Definiciones.

En estadística, el análisis de la regresión es un proceso estadístico para estimar las relaciones entre variables. Incluye muchas técnicas para el modelado y análisis de diversas variables, cuando la atención se centra en la relación entre una variable dependiente y una o más variables independientes (o predictoras). Más específicamente, el análisis de regresión ayuda a entender cómo el valor de la variable dependiente varía al cambiar el valor de una de las variables independientes, manteniendo el valor de las otras variables independientes fijas. Más comúnmente, el análisis de regresión estima la esperanza condicional de la variable dependiente dadas las variables independientes - es decir, el valor promedio de la variable dependiente cuando se fijan las variables independientes. Con menor frecuencia, la atención se centra en un cuantil, u otro parámetro de localización de la distribución condicional de la variable dependiente dadas las variables independientes. En todos los casos, el objetivo de la estimación es una función de las variables independientes llamada la función de regresión. En el análisis de regresión, también es de interés caracterizar la variación de la variable dependiente en torno a la función de regresión, la cual puede ser descrita por una distribución de probabilidad.

Se denomina correlación al vínculo recíproco o correspondiente que existe entre dos o más elementos. El concepto se emplea de diferentes maneras de acuerdo al contexto.

Correlación en el ámbito de las matemáticas y las estadísticas, la correlación alude a la proporcionalidad y la relación lineal que existe entre distintas variables. Si los valores de una variable se modifican de manera sistemática con respecto a los valores de otra, se dice que ambas variables se encuentran correlacionadas.

1.4.2. Curva de regresión y coeficiente de determinación.

En estadística la regresión lineal o ajuste lineal es un modelo matemático usado para aproximar la relación de dependencia entre una variable dependiente Y, las variables independientes Xi y un término aleatorio.

En estadística, el coeficiente de determinación, denominado R^2 y pronunciado R cuadrado, es un estadístico usado en el contexto de un modelo estadístico cuyo principal propósito es predecir futuros resultados o probar una hipótesis. El coeficiente determina la calidad del modelo para replicar los resultados, y la proporción de variación de los resultados que puede explicarse por el modelo.1 Hay varias definiciones diferentes para R^2 que son algunas veces equivalentes. Las más comunes se refieren a la regresión lineal. En este caso, el R^2 es simplemente el cuadrado del coeficiente de correlación de Pearson, lo cual es sólo cierto para la regresión lineal simple. Si existen varios resultados para una única variable, es decir, para una X existe una Y, Z... el coeficiente de determinación resulta del cuadrado del coeficiente de determinación múltiple. En ambos casos el R^2 adquiere valores entre 0 y 1. Existen casos dentro de la definición computacional de R^2 donde este valor puede tomar valores negativos.

1.4.3. Regresión y correlación lineal.

La correlación lineal y la regresión lineal simple son métodos estadísticos que estudian la relación lineal existente entre dos variables. Antes de profundizar en cada uno de ellos, conviene destacar algunas diferencias:

- La correlación cuantifica como de relacionadas están dos variables, mientras que la regresión lineal consiste en generar una ecuación (modelo) que, basándose en la relación existente entre ambas variables, permita predecir el valor de una a partir de la otra.
- El cálculo de la correlación entre dos variables es independiente del orden o asignación de cada variable a X e Y, mide únicamente la relación entre ambas sin considerar dependencias. En el caso de la regresión lineal, el modelo varía según qué variable se considere dependiente de la otra (lo cual no implica causa-efecto).

A nivel experimental, la correlación se suele emplear cuando ninguna de las variables se ha controlado, simplemente se han medido ambas y se desea saber si están relacionadas. En el caso de estudios de regresión lineal, es más común que una de las variables se controle (tiempo, concentración de reactivo, temperatura…) y se mida la otra.

Por norma general, los estudios de correlación lineal preceden a la generación de modelos de regresión lineal. Primero se analiza si ambas variables están correlacionadas y, en caso de estarlo, se procede a generar el modelo de regresión.

Para estudiar la relación lineal existente entre dos variables continuas es necesario disponer de parámetros que permitan cuantificar dicha relación. Uno de estos parámetros es la covarianza, que indica el grado de variación conjunta de dos variables aleatorias.

La covarianza depende de las escalas en que se miden las variables estudiadas, por lo tanto, no es comparable entre distintos pares de variables. Para poder hacer comparaciones se estandariza la covarianza, generando lo que se conoce como coeficientes de correlación. Existen diferentes tipos, de entre los que destacan el coeficiente de Pearson, Rho de Spearman y Tau de Kendall.

1.4.4. Otros tipos de regresión.

Podemos clasificar los tipos de regresión según diversos criterios.

En primer lugar, en función del número de variables independientes:

- Regresión simple: Cuando la variable Y depende únicamente de una única variable X.
- Regresión múltiple: Cuando la variable Y depende de varias variables (X1, X2, ..., Xr)

En segundo lugar, en función del tipo de función f(X):

- Regresión lineal: Cuando f(X) es una función lineal.
- Regresión no lineal: Cuando f(X) no es una función lineal.

En tercer lugar, en función de la naturaleza de la relación que exista entre las dos variables:

- La variable X puede ser la causa del valor de la variable Y.

Por ejemplo, en toxicología, si X = Dosis de la droga e Y = Mortalidad, la mortalidad se atribuye a la dosis administrada y no a otras causas.

Puede haber simplemente relación entre las dos variables.

Por ejemplo, en un estudio de medicina en que se estudian las variables X = Peso e Y = Altura de un grupo de individuos, puede haber relación entre las dos, aunque difícilmente una pueda considerarse causa de la otra.

1.4.5 Análisis de atributos.

Su principal objetivo es el de evitar un error muy común consistente en tratar de encontrar la forma de mejorar un producto, servicio o proceso analizándolo como un todo. Muchas veces, la búsqueda de una idea global, salvadora, que mejore el todo, impide descubrir la característica específica que, por sí sola, podría producir el resultado deseado.

Características para las Gráficas de Control de Atributos

- Están basadas en decisiones de pasa/no pasa.
- Se pueden aplicar en casi cualquier operación donde se recolectan datos.
- Se utilizan en características de calidad que no pueden ser medidas o que son costosas o difíciles de medir.

A diferencia de las gráficas de control de datos variables, las gráficas de datos atributos se pueden establecer para una característica de calidad o para muchas.

Tipos de Gráficas de Atributos:

- Defectivos
  – np - número de unidades no-conformes
  – p - proporción de unidades no-conformes
- Defectos
  – c - número de defectos
  – u - proporción de defectos

UNIDAD II

CALCULO DE PROBABILIDADES

2.1. Introducción al cálculo de probabilidades.

En la vida cotidiana aparecen muchas situaciones en las que los resultados observados son diferentes, aunque las condiciones iniciales en las que se produce la experiencia sean las mismas. Por ejemplo, al lanzar una moneda unas veces resultará cara y otras, cruz. Estos fenómenos, denominados aleatorios, se ven afectados por la incertidumbre.

En el lenguaje habitual, frases como "probablemente...", "es poco probable que...", "hay muchas posibilidades de que..." hacen referencia a esta incertidumbre.

La teoría de la probabilidad pretende ser una herramienta para modelizar y tratar con situaciones de este tipo. Por otra parte, cuando aplicamos las técnicas estadísticas a la recogida, análisis e interpretación de los datos, la teoría de la probabilidad proporciona una base para evaluar la fiabilidad de las conclusiones alcanzadas y las inferencias realizadas.

El objetivo del Cálculo de Probabilidades es el estudio de métodos de análisis del comportamiento de fenómenos aleatorios.

2.1.1 La medida de probabilidad. Espacio Probabilístico.

Para medir la incertidumbre existente en un experimento aleatorio1 dado, se parte de un espacio muestral M en el que se incluyen todos los posibles resultados individuales del experimento (sucesos elementales); es decir, el conjunto muestral es un conjunto exhaustivo (contiene todas las posibles ocurrencias) y mutuamente exclusivo (no pueden darse dos ocurrencias a la vez). Una vez definido el espacio muestral, el objetivo consiste en asignar a todo suceso compuesto $A \subset M$ un número real que mida el grado de incertidumbre sobre su ocurrencia. Para obtener medidas con significado matemático claro y practico, se imponen ciertas propiedades intuitivas que definen una clase de medidas que se conocen como medidas de probabilidad.

Definición Medida de Probabilidad. Una función p que proyecta los subconjuntos $A \subset M$ en el intervalo [0, 1] se llama medida de probabilidad si satisface los siguientes axiomas:

Axioma 1: Un experimento se denomina aleatorio cuando puede dar resultados distintos al realizarse en las mismas condiciones (por ejemplo, lanzar un dado al aire y observar el número resultante).

Formalmente, una medida de probabilidad se define sobre una σ-´algebra del espacio muestral, que es una colección de subconjuntos que es cerrada para los operadores de unión $A \cup B$ y complementario $\bar{A} = M \backslash A$ (por tanto, también para intersecciones $A \cap B = \bar{A} \cup \bar{B}$).

Sin embargo, optamos por una definición menos rigurosa y más intuitiva para introducir este concepto.

Axioma 2: Para cualquier sucesión infinita, $A1, A2, \ldots$, de subconjuntos disjuntos de M, se cumple la igualdad.

El Axioma 1 establece que, independientemente de nuestro grado de certeza, ocurrirá un elemento del espacio muestral M (es decir, el conjunto M es exhaustivo). El Axioma 2 es una fórmula de agregación que se usa para calcular la probabilidad de la unión de subconjuntos disjuntos. Establece que la incertidumbre de un cierto subconjunto es la suma de las incertidumbres de sus partes (disjuntas). Nótese que esta propiedad también se cumple para sucesiones finitas.

En teoría de probabilidades, un espacio probabilístico o espacio de probabilidad es un concepto matemático que sirve para modelar un cierto experimento aleatorio.

En general un espacio probabilístico está integrado por tres componentes. Primero, el conjunto (llamado espacio muestral) de los posibles resultados del experimento, llamados sucesos elementales. Segundo, por la colección de todos los sucesos aleatorios (no solo los elementales), que es una σ-álgebra sobre. El par es lo que se conoce como un espacio de medida. Por último, una medida de probabilidad o función de probabilidad, que asigna una probabilidad a todo suceso y que verifica los llamados axiomas de Kolmogórov.

2.1.2. Probabilidad condicionada.

Probabilidad condicional es la probabilidad de que ocurra un evento A, sabiendo que también sucede otro evento B. La probabilidad condicional se escribe P(A|B) o P(A/B), y se lee «la probabilidad de A dado B».

No tiene por qué haber una relación causal o temporal entre A y B. A puede preceder en el tiempo a B, sucederlo o pueden ocurrir simultáneamente. A puede causar B, viceversa o pueden no tener relación causal. Las relaciones causales o temporales son nociones que no pertenecen al ámbito de la probabilidad. Pueden desempeñar un papel o no, dependiendo de la interpretación que se le dé a los eventos.

Un ejemplo clásico es el lanzamiento de una moneda para luego lanzar un dado. ¿Cuál es la probabilidad que en el dado salga un 6 dado que ya haya salido una cara en la moneda? Esta probabilidad se denota de esta manera: P(6|C).

El condicionamiento de probabilidades puede lograrse aplicando el teorema de Bayes.

La falacia de la probabilidad condicional se basa en asumir que P(A|B) es casi igual a P(B|A). El matemático John Allen Paulos analiza en su libro El hombre anumérico este error muy común cometido por personas que desconocen la probabilidad.

Como la probabilidad está ligada a nuestra ignorancia sobre los resultados de la experiencia, el hecho de que ocurra un suceso, puede cambiar la probabilidad de los demás. El proceso de realizar la historia clínica, explorar y realizar pruebas complementarias ilustra este principio.

La probabilidad condicional es la probabilidad de algún evento A , dada la ocurrencia de algún otro evento B . Esto está denotado por P (A | B) y se lee "la probabilidad de A , dado B ". En otras palabras, estamos calculando probabilidades condicionales al conocer información adicional parcialmente a través del experimento.

2.1.3. Teoremas asociados.

El teorema de Bayes, en la teoría de la probabilidad, es una proposición planteada por el matemático inglés Thomas Bayes (1702-1761)1 y publicada póstumamente en 1763,2 que expresa la probabilidad condicional de un evento aleatorio A dado B en términos de la distribución de probabilidad condicional del evento B dado A y la distribución de probabilidad marginal de solo A.

En términos más generales y menos matemáticos, el teorema de Bayes es de enorme relevancia puesto que vincula la probabilidad de A dado B con la probabilidad de B dado A. Es decir, por ejemplo, que sabiendo la probabilidad de tener un dolor de cabeza dado que se tiene gripe, se podría saber (si se tiene algún dato más), la probabilidad de tener gripe si se tiene un dolor de cabeza. Muestra este sencillo ejemplo la alta relevancia del teorema en cuestión para la ciencia en todas sus ramas, puesto que tiene vinculación íntima con la comprensión de la probabilidad de aspectos causales dados los efectos observados.

2.2. Variable aleatoria.

Una variable aleatoria es una función que asigna un valor, usualmente numérico, al resultado de un experimento aleatorio. Por ejemplo, los posibles resultados de tirar un dado dos veces: (1, 1), (1, 2), etc. o un número real (p.e., la temperatura máxima medida a lo largo del día en una ciudad concreta).

Los valores posibles de una variable aleatoria pueden representar los posibles resultados de un experimento aún no realizado, o los posibles valores de una cantidad cuyo valor actualmente existente es incierto (p.e., como resultado de medición incompleta o imprecisa). Intuitivamente, una variable aleatoria puede tomarse como una cantidad cuyo valor no es fijo pero puede tomar diferentes valores; una distribución de probabilidad se usa para describir la probabilidad de que se den los diferentes valores. En términos formales una variable aleatoria es una función definida sobre un espacio de probabilidad.

Las variables aleatorias suelen tomar valores reales, pero se pueden considerar valores aleatorios como valores lógicos, funciones o cualquier tipo de elementos (de un espacio medible). El término elemento aleatorio se utiliza para englobar todo ese tipo de conceptos relacionados. Un concepto relacionado es el de proceso estocástico, un conjunto de variables aleatorias ordenadas (habitualmente por orden o tiempo).

Una variable aleatoria puede concebirse como un valor numérico que está afectado por el azar. Dada una variable aleatoria no es posible conocer con certeza el valor que tomará esta al ser medida o determinada, aunque sí se conoce que existe una distribución de probabilidad asociada al conjunto de valores posibles. Por ejemplo, en una epidemia de cólera, se sabe que una persona cualquiera puede enfermar o no (suceso), pero no se sabe cuál de los dos sucesos va a ocurrir. Solamente se puede decir que existe una probabilidad de que la persona enferme.

Para trabajar de manera sólida con variables aleatorias en general es necesario considerar un gran número de experimentos aleatorios, para su tratamiento estadístico, cuantificar los resultados de modo que se asigne un número real a cada uno de los resultados posibles del experimento. De este modo se establece una relación funcional entre elementos del espacio muestral asociado al experimento y números reales.

2.2.1. Concepto de variable aleatoria. Probabilidad inducida.

Una variable es un símbolo que actúa en las funciones, las fórmulas, los algoritmos y las proposiciones de las matemáticas y la estadística. Según sus características, las variables se clasifican de distinto modo.

Variable aleatoria

Se denomina variable aleatoria (o estocástica) a la función que adjudica eventos posibles a números reales (cifras), cuyos valores se miden en experimentos de tipo aleatorio. Estos valores posibles representan los resultados de experimentos que todavía no se llevaron a cabo o cantidades inciertas.

Cabe destacar que los experimentos aleatorios son aquellos que, desarrollados bajo las mismas condiciones, pueden ofrecer resultados diferentes. Arrojar una moneda al aire para ver si sale cara o ceca es un experimento de este tipo.

La variable aleatoria, en definitiva, permite ofrecer una descripción de la probabilidad de que se adoptan ciertos valores. No se sabe de manera precisa qué valor adoptará la variable cuando sea determinada o medida, pero sí se puede conocer cómo se distribuyen las probabilidades vinculadas a los valores posibles. En dicha distribución incide el azar.

Tal como hemos comentado, la definición formal de variable aleatoria impone una restricción matemática en la formulación vista hasta el momento.

Definiremos una variable aleatoria como una aplicación de Ω en el conjunto de números reales, s decir, para todo número real x, el conjunto de resultados elementales tales que la variable aleatoria toma sobre ellos valores inferiores o iguales a x ha de ser un suceso sobre el cual podamos definir una probabilidad.

Dicha propiedad recibe el nombre de medibilidad y por tanto podríamos decir que una variable aleatoria es una función medible de Ω en los reales.

Esta condición nos asegura que podremos calcular sin problemas, probabilidades sobre intervalos de la recta real a partir de las probabilidades de los sucesos correspondientes.

La expresión anterior se leería de la manera siguiente: La probabilidad de que la variable aleatoria tome valores inferiores o iguales a x es igual a la probabilidad del suceso formado por el conjunto de resultados elementales sobre los que el valor de la variable es menor o igual que x.

La probabilidad obtenida de esta manera se denomina probabilidad inducida.

Se puede comprobar que, a partir de la condición requerida, se pueden obtener probabilidades sobre cualquier tipo de intervalo de la recta real.

2.2.2. Función de distribución.

En la teoría de la probabilidad y en estadística, la Función de Distribución Acumulada (FDA, designada también a veces simplemente como FD) o función de probabilidad acumulada asociada a una variable aleatoria real: X (mayúscula) sujeta a cierta ley de distribución de probabilidad, es una función matemática de la variable real: x (minúscula); que describe la probabilidad de que X tenga un valor menor o igual que x .

Intuitivamente, asumiendo la función f como la ley de distribución de probabilidad, la FDA sería la función con la recta real como dominio, con imagen del área hasta aquí de la función f, siendo aquí el valor x para la variable aleatoria real X.

La FDA asocia a cada valor x, la probabilidad del evento: "la variable X toma valores menores o iguales a x".

El concepto de FDA puede generalizarse para modelar variables aleatorias multivariantes.

2.2.3. Variables aleatorias discretas y continuas.

Se denomina variable aleatoria discreta aquella que sólo puede tomar un número finito de valores dentro de un intervalo. Por ejemplo, el número de componentes de una manada de lobos, puede ser 4 ó 5 ó 6 individuos, pero nunca 5,75 ó 5,87. Otros ejemplos de variable discreta serían el número de pollos de gorrión que llegan a volar del nido o el sexo de los componentes de un grupo familiar de babuinos.

Toda variable aleatoria discreta tiene asociada una función de probabilidad, que, a cada valor, le marca la probabilidad de que la variable tome dicho valor. Esta probabilidad viene a jugar el mismo papel que la frecuencia relativa en los temas de estadística.

Una variable aleatoria continua es una función X que asigna a cada resultado posible de un experimento un número real. Si X puede asumir cualquier valor en algún intervalo I (el intervalo puede ser acotado o desacotado), se llama una variable aleatoria continua. Si puede asumir solo varios valores distintos, se llama una variable aleatoria discreta.

Una variable aleatoria X es continua si su función de distribución es una función continua.

En la práctica, se corresponden con variables asociadas con experimentos en los cuales la variable medida puede tomar cualquier valor en un intervalo: mediciones biométricas, intervalos de tiempo, áreas, etc.

Dentro de las variables aleatorias continuas tenemos las variables aleatorias absolutamente continuas.

Una variable aleatoria con distribución absolutamente continua, por extensión, se clasifica como variable aleatoria absolutamente continua.

En el presente manual, todas las variables aleatorias continuas con las que trabajemos pertenecen al grupo de las variables absolutamente continuas, en particular, los ejemplos y casos expuestos.

2.3. Características de una variable.

Las variables como entidades empíricas del problema de investigación presentan un conjunto de características significativas tales como:

- Están contenidas esencialmente en el título, el problema, el objetivo y las respectivas hipótesis de la investigación. En virtud de ello es que no se puede agregar nuevas variables de las que ya existen en los ítems mencionados.
- Son aspectos que cambian o adoptan distintos valores. Esto significa que las variables al ser medidas y observadas expresan diferencias entre los rasgos, cualidades y atributos de las unidades de análisis.
- Son enunciados que expresan rasgos característicos de los problemas medibles empíricamente. Estas variables en la práctica social pueden ser medidas y observadas con instrumentos convencionales, en mérito de que contienen rasgos, propiedades y cualidades.
- Son susceptibles de descomposición empírica. Dicho de otro término, que las variables pueden desagregarse en indicadores, índices, subíndices e ítems.

2.3.1. Esperanza de una variable aleatoria.

En estadística la esperanza matemática (también llamada esperanza, valor esperado, media poblacional o media) de una variable aleatoria, es el número que formaliza la idea de valor medio de un fenómeno aleatorio.

Cuando la variable aleatoria es discreta, la esperanza es igual a la suma de la probabilidad de cada posible suceso aleatorio multiplicado por el valor de dicho suceso. Por lo tanto, representa la cantidad media que se "espera" como resultado de un experimento aleatorio cuando la probabilidad de cada suceso se mantiene constante y el experimento se repite un elevado número de veces. Cabe decir que el valor que toma la esperanza matemática en algunos casos puede no ser "esperado" en el sentido más general de la palabra (el valor de la esperanza puede ser improbable o incluso imposible).

La esperanza matemática de una variable aleatoria es una característica numérica que proporciona una idea de la localización de la variable aleatoria sobre la recta real. Decimos que es un parámetro de centralización o de localización.

Su interpretación intuitiva o significado se corresponde con el valor medio teórico de los posibles valores que pueda tomar la variable aleatoria, o también con el centro de gravedad de los valores de la variable supuesto que cada valor tuviera una masa proporcional a la función de densidad en ellos.

La definición matemática de la esperanza en el caso de las variables aleatorias discretas se corresponde directamente con las interpretaciones proporcionadas en el párrafo anterior.

En caso de que el recorrido sea infinito la esperanza existe si la serie resultante es absolutamente convergente, condición que no siempre se cumple.

La definición se corresponde con un promedio ponderado según su probabilidad de los valores del recorrido y, por tanto, se corresponde con la idea de un valor medio teórico.

2.3.2. Momentos de una variable aleatoria.

Cuando la distribución de probabilidad de una variable aleatoria no es conocida, diversas características de ella pueden proporcionar una descripción general de la misma.

Entre las distintas características de una distribución ocupan un importante lugar los momentos, entre los que cabe destacar los diferentes tipos que definimos a continuación:

Momentos no centrados

Momentos centrados en media

Los momentos centrados se calculan, como los no centrados, teniendo en cuenta la definición de esperanza de una función de una variable aleatoria.

La varianza de una variable, si existe, es el valor medio de las dispersiones cuadráticas de los valores de la variable respecto de su media. Por este motivo, tanto la varianza como su raíz cuadrada, σX, que se denomina desviación típica, se usan, como se verá posteriormente, como medidas de la dispersión de la variable.

2.3.4. Funciones asociadas a una variable aleatoria.

Una función que asocia un número real, perfectamente definido, a cada punto muestral.

A veces las variables aleatorias (v.a.) están ya implícitas en los puntos muéstrales.

a función que caracteriza las variables continuas es aquella función f positiva e integrable en los reales, tal que acumulada desde $-\infty$ hasta un punto x, nos proporciona el valor de la función de distribución en x, F(x). Recibe el nombre de función de densidad de la variable aleatoria continua.

Las funciones de densidad discreta y continua tienen, por tanto, un significado análogo, ambas son las funciones que acumuladas (en forma de sumatorio en el caso discreto o en forma de integral en el caso continuo) dan como resultado la función de distribución.

La diferencia entre ambas, sin embargo, es notable.

La función de densidad discreta toma valores positivos únicamente en los puntos del recorrido y se interpreta como la probabilidad de la que la variable tome ese valor $f(x) = P(X = x)$.

La función de densidad continua toma valores en el conjunto de números reales y no se interpreta como una probabilidad. No está acotada por 1, puede tomar cualquier valor positivo. Es más, en una variable continua se cumple que probabilidades definidas sobre puntos concretos siempre son nulas.

2.4. Modelos de los de distribución de probabilidad.

Una distribución de probabilidad queda definida y caracterizada por:

1.- la especificación de la variable aleatoria y su campo de variación.

2.- la especificación de su asignación de probabilidades, mediante la función de distribución. (Alternativamente mediante la f.cuantía o densidad, la F.C. o la F.G.M.(si existe).(Estas son las FUNCIONES DE DEFINICIÓN).

Si un conjunto dado de distribuciones tiene sus funciones de distribución con la misma ESTRUCTURA FUNCIONAL, diremos que pertenece a la misma FAMILIA DE DISTRIBUCIONES, al mismo MODELO DE PROBABILIDAD o a la misma DISTRIBUCIÓN-TIPO.

La estructura matemática de las funciones de definición que caracterizan un modelo de probabilidad suelen depender de uno o más parámetros.Estos parámetros son los PARÁMETROS DE LA DISTRIBUCIÓN(TIPO), y tienen un importancia fundamental, en Estadística matemática y sobre todo en INFERENCIA ESTADÍSTICA.

MODELOS DISCRETOS

Aunque en adelante hablemos de distribución "tal", nos estaremos refiriendo al modelo tal.

Los modelos discretos, son modelos de probabilidad de variable aleatoria discreta. Los más importante son los modelos de BERNOUILLI (especialmente "la distribución binomial") y la "distribución de Poisson".

DISTRIBUCIÓN DICOTÓMICA. (Bernoulli).

Si una variable aleatoria X sigue o tiene una distribución dicotómica de parámetro p se expresa como X ~ D(p).

Modeliza situaciones en las que:

- Se realiza una prueba
- Que sólo puede dar dos resultados posibles: A y A
- La probabilidad del resultado A es P(A) = p y la del resultado A es P(A)= q=1-p.
- En estas circunstancias la variable aleatoria X significa "nº de resultados A que se obtienen.

DISTRIBUCIÓN BINOMIAL.

Si una variable aleatoria, X, sigue una distribución binomial de parámetros n y p se expresa como: X ~ B(n,p).

Situaciones que modeliza:

Se realiza un número n de pruebas (separadas o separables).

Cada prueba puede dar dos únicos resultados A y Ã

La probabilidad de obtener un resultado A es p y la de obtener un resultado Ã es q, con q= 1-p, en todas las pruebas. Esto implica que las pruebas se realizan exactamente en las mismas condiciones. Si se trata de extracciones, (muestreo), las extracciones deberán ser con devolución (reemplazamiento) (M.A.S).

Es fácil comprobar que considerando estas condiciones la función de cuantía de la variable es precisamente la que se ha especificado arriba.

La función de distribución quedará como F(x) = S P(x), sin una expresión analítica concreta.

Los indicadores-momentos (media y varianza) pueden obtenerse a partir de la función de cuantía (operador esperanza) o a a partir de F.G.M.

DISTRIBUCIÓN HIPERGEOMÉTRICA

Dada la siguiente situación:

Una población constituida por N individuos en total.

De los cuales Np individuos son del tipo A, y Nq individuos son del tipo Ã.

De forma que la proporción de individuos A que hay en la población es p, y la proporción de individuos de tipo Ã, es q (p+q=1).

Se realizan n (pruebas) extracciones sin reemplazamiento.

De forma que la probabilidad de extraer un individuo A (Ã) en una de las extracciones depende de los resultados de las pruebas anteriores.

Si consideramos la variable aleatoria X = nº de resultados A obtenidos en las n extracciones, X seguirá una distribución hipergeométrica. X~H(N,n,p).

La distribución hipergeométrica es semejante a la binomial, excepto en el hecho de que las pruebas no mantienen constantes las probabilidades de A y Ã.

- La media de la distribución hipergeométrica es m = np
- La varianza de la distribución es s2 = npq (N-1/(N-n))

DISTRIBUCIÓN DE POISSON

Formalmente: dada una variable aleatoria X con campo de variación.

Situaciones que modeliza:

Se observa la ocurrencia de hechos de cierto tipo durante un período de tiempo o a lo largo de un espacio, considerados unitarios.

El tiempo (o el espacio) pueden considerarse homogéneos, respecto al tipo de hechos estudiados, al menos durante el período experimental; es decir, que no hay razones para suponer que en ciertos momentos los hechos sean más probables que otros.

En un instante (infinitesimal) sólo puede producirse como mucho un hecho (se podrá producir o uno o ninguno).

La probabilidad de que se produzca un hecho en un intervalo infinitesimal es prácticamente proporcional a la amplitud del intervalo infinitesimal.

Si en estas circunstancias la variable aleatoria X = nº de hechos que se producen en un intervalo unitario sigue una distribución de Poisson, que cómo veremos tendrá por parámetro l el número medio de hechos que pueden producirse en el intervalo unitario.

MODELOS CONTINUOS

DISTRIBUCIÓN UNIFORME (DE V. CONTINUA)

Dada una variable aleatoria continua, X , definida en el intervalo [a,b] de la recta real, diremos que X tiene una distribución uniforme en el intervalo [a,b] cuando su función de densidad sea: X ~ U([a,b])

f(x)= 1/(b-a) para x Î [a,b].

DISTRIBUCIÓN EXPONENCIAL

Dada una variable aleatoria continua, X , definida para valores reales positivos. Diremos que X tiene una distribución exponencial de parámetro a cuando su función de densidad sea: f(x) = a e-a x para x ³ 0 (siendo el parámetro a positivo).

DISTRIBUCIÓN NORMAL

La distribución normal es la más importante de todas las distribuciones de probabilidad. Es una distribución de variable continua con campo de variación [-¥ ,¥], que queda especificada a través de dos parámetros (que acaban siendo la media y la desviación típica de la distribución).

Importancia de la distribución Normal.

a) Enorme número de fenómenos que puede modelizar: Casi todas las características cuantitativas de las poblaciones muy grades tienden a aproximar su distribución a una distribución normal.

b) Muchas de las demás distribuciones de uso frecuente, tienden a distribuirse según una Normal, bajo ciertas condiciones.

c) (En virtud del teorema central del límite). Todas aquellas variables que pueden considerarse causadas por un gran número de pequeños efectos (como pueden ser los errores de medida) tienden a distribuirse según una distribución normal.

La probabilidad de cualquier intervalo se calcularía integrando la función de densidad a lo largo de ese de intervalo, pero no es necesario nunca resolver la integral pues existen tablas que nos evitan este problema.

2.4.1. Distribuciones Binomial y Poisson.

En estadística, la distribución binomial es una distribución de probabilidad discreta que cuenta el número de éxitos en una secuencia de n ensayos de Bernoulli independientes entre sí, con una probabilidad fija p de ocurrencia del éxito entre los ensayos. Un experimento de Bernoulli se caracteriza por ser dicotómico, esto es, solo dos resultados son posibles. A uno de estos se denomina éxito y tiene una probabilidad de ocurrencia p y al otro, fracaso, con una probabilidad2 q=1-p. En la distribución binomial el anterior experimento se repite n veces, de forma independiente, y se trata de calcular la probabilidad de un determinado número de éxitos. Para n = 1, la binomial se convierte, de hecho, en una distribución de Bernoulli.

Existen muchas situaciones en las que se presenta una experiencia binomial. Cada uno de los experimentos es independiente de los restantes (la probabilidad del resultado de un experimento no depende del resultado del resto). El resultado de cada experimento ha de admitir sólo dos categorías (a las que se denomina éxito y fracaso). El valor de ambas posibilidades ha de ser constante en todos los experimentos, y se denotan como p y q respectivamente, o p y 1-p de forma alternativa.

Se designa por X a la variable que mide el número de éxitos que se han producido en los n experimentos.

Cuando se dan estas circunstancias, se dice que la variable X sigue una distribución de probabilidad binomial, y se denota B(n,p).

Una distribución binomial es una distribución de probabilidad discreta que describe el número de éxitos al realizar n experimentos independientes entre sí, acerca de una variable aleatoria.

Existen una gran diversidad de experimentos o sucesos que pueden ser caracterizados bajo esta distribución de probabilidad. Imaginemos el lanzamiento de una moneda en el que definimos el suceso “sacar cara” como el éxito. Si lanzamos 5 veces la moneda y contamos los éxitos (sacar cara) que obtenemos, nuestra distribución de probabilidades se ajustaría a una distribución binomial.

Por lo tanto, la distribución binomial se entiende como una serie de pruebas o ensayos en la que solo podemos tener 2 resultados (éxito o fracaso), siendo el éxito nuestra variable aleatoria.

Propiedades de la distribución binomial

Para que una variable aleatoria se considere que sigue una distribución binomial, tiene que cumplir las siguientes propiedades:

En cada ensayo, experimento o prueba solo son posibles dos resultados (éxito o fracaso).

La probabilidad del éxito ha de ser constante. Esta se representa mediante la letra p. La probabilidad de que salga cara al lanzar una moneda es 0,5 y esta es constante dado que la moneda no cambia en cada experimento y las probabilidades de sacar cara es constate.

La probabilidad de fracaso ha de ser también constate. Esta se representa mediante la letra q = 1-p. Es importante fijarse que mediante esa ecuación, sabiendo p o sabiendo q, podemos obtener la que nos falte.

El resultado obtenido en cada experimento es independiente del anterior. Por lo tanto lo que ocurra en cada experimento no afecta a los siguientes.

Los sucesos son mutuamente excluyentes, es decir, no pueden ocurrir los 2 al mismo tiempo. No se puede ser hombre y mujer al mismo tiempo o que al lanzar una moneda salga cara y cruz al mismo tiempo.

Los sucesos son colectivamente exhaustivos, es decir, al menos uno de los 2 ha de ocurrir. Si no se es hombre, se es mujer y si se lanza una moneda, si no sale cara ha de salir cruz.

La variable aleatoria que sigue una distribución binomial se suele representar como X~(n,p). n representa el número de ensayos o experimentos y p la probabilidad de éxito.

La Distribución binomial es uno de los modelos de distribución teórica de probabilidad que se utiliza cuando la variable aleatoria discreta es el número de éxitos en una muestra compuesta por n observaciones. Es una de las distribuciones de probabilidad más útiles que se emplea en control de calidad, producción, investigaciones, etc.

La Distribución Binomial está relacionada con la distribución de Bernoulli que es una distribución de variable aleatoria X que toma solamente valores de cero y uno (éxito y fracaso), cuando se realiza un solo experimento.

La distribución binomial se aplica cuando se realizan un número "n" de veces el experimento de Bernouilli, siendo cada ensayo independiente del anterior. La variable puede tomar valores entre 0 (si todos los experimentos han sido fracaso) y n (si todos los experimentos han sido éxitos).

La distribución binomial es una distribución de probabilidades que surge al cumplirse las siguientes condiciones:

- El número de ensayos o repeticiones del experimento (n) es constante.
- En cada ensayo hay sólo dos posibles resultados (éxito o fracaso, defectuoso o noi defectuoso).
- La probabilidad de cada resultado posible en cualquier ensayo permanece constante.
- En cada ensayo, los dos resultados posibles son mutuamente excluyentes.
- Los resultados de cada ensayo son independientes entre si.

2.4.2. Otras distribuciones discretas.

En teoría de la probabilidad y estadística, la distribución de probabilidad de una variable aleatoria es una función que asigna a cada suceso definido sobre la variable la probabilidad de que dicho suceso ocurra. La distribución de probabilidad está definida sobre el conjunto de todos los sucesos y cada uno de los sucesos es el rango de valores de la variable aleatoria. También puede decirse que tiene una relación estrecha con las distribuciones de frecuencia. De hecho, una distribución de probabilidades puede comprenderse como una frecuencia teórica, ya que describe cómo se espera que varíen los resultados.

La distribución de probabilidad está completamente especificada por la función de distribución, cuyo valor en cada x real es la probabilidad de que la variable aleatoria sea menor o igual que x.

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.

Cuando se analiza un experimento aleatorio, se descubren factores de comportamiento de la probabilidad que siguen modelos propios y distintivos. Por ello, es frecuente asociar a estos experimentos una «función de probabilidad», que puede adoptar diversas formas y regirse por principios diferentes y cuyo estudio arroja luz sobre la naturaleza y las características del fenómeno físico o social ligado al experimento.

Función de distribución. -

La distribución Normal suele conocerse como la "campana de gauss".

En teoría de la probabilidad y estadística, la distribución de probabilidad de una variable aleatoria es una función que asigna a cada suceso definido sobre la variable aleatoria la probabilidad de que dicho suceso ocurra. La distribución de probabilidad está definida sobre el conjunto de todos los eventos rango de valores de la variable aleatoria.

Cuando la variable aleatoria toma valores en el conjunto de los números reales, la distribución de probabilidad está completamente especificada por la función de distribución, cuyo valor en cada real x es la probabilidad de que la variable aleatoria sea menor o igual que x.

La distribución Uniforme discreta

Tenemos esta distribución cuando el resultado de una experiencia aleatoria puede ser un conjunto finito de n posibles resultados, todos ellos igualmente probables.

La distribución Hipergeométrica

Este modelo presenta similitudes con el Binomial, pero sin la suposición de independencia de éste último. Veámoslo:

Partimos de un conjunto formado por N individuos divididos en dos categorías mutuamente excluyentes: A y Ac; de manera que N1 individuos pertenecen a la categoría A y N2 individuos, a la categoría Ac.

La dependencia se debe al hecho de que N es finito y las extracciones se efectúan sin reemplazamiento. El caso de extracciones con reemplazamiento sería equivalente al de N infinito y se resolvería mediante el modelo Binomial.

La distribución Geométrica o de Pascal

Definamos una experiencia aleatoria cuyo resultado sólo puede ser el suceso A o su complementario Ac, y que se repite secuencialmente hasta que aparece el suceso A por primera vez.

Definamos la variable aleatoria X como el número de veces que repetimos la experiencia en condiciones independientes hasta que se dé A por primera vez. Bajo estas condiciones, decimos que la variable X sigue una distribución geométrica o de Pascal de parámetro p = P(A).

Algunas puntualizaciones de la definición de X:

- Notése que, en esta definición, condiciones independientes significa que p, la probabilidad de A, y 1 − p, la de su complementario Ac, no varían a lo largo de las sucesivas repeticiones de la experiencia.
- Tal y como la hemos definido, X se refiere al número de lanzamientos hasta que se produce A, pero sin contabilizar el último caso en que se da A. Por dicha razón X puede tomar los valores k = 0, 1, 2, ... con probabilidad no nula.

La distribución Binomial negativa

Puede definirse como una generalización del modelo Geométrico o de Pascal. Así, dado un suceso A y su complementario Ac, cuando X representa el número de veces que se da Ac (ausencias, fallos, etc.) hasta que se produce r veces el suceso A, en una serie de repeticiones de la experiencia aleatoria en condiciones independientes, decimos que X sigue la distribución Binomial negativa. Nótese que, cuando r = 1, tenemos exactamente el modelo geométrico.

Este modelo queda definido por dos parámetros p (la probabilidad de A: p = P(A)) y r (el número de veces que debe producirse A para que detengamos la experiencia).

2.4.3. Distribución normal.

La distribución normal es una distribución con forma de campana donde las desviaciones estándar sucesivas con respecto a la media establecen valores de referencia para estimar el porcentaje de observaciones de los datos. Estos valores de referencia son la base de muchas pruebas de hipótesis, como las pruebas Z y t.

En estadística y probabilidad se llama distribución normal, distribución de Gauss, distribución gaussiana o distribución de Laplace-Gauss, a una de las distribuciones de probabilidad de variable continua que con más frecuencia aparece en estadística y en la teoría de probabilidades.

La gráfica de su función de densidad tiene una forma acampanada y es simétrica respecto de un determinado parámetro estadístico. Esta curva se conoce como campana de Gauss y es el gráfico de una función gaussiana.

La importancia de esta distribución radica en que permite modelar numerosos fenómenos naturales, sociales y psicológicos. Mientras que los mecanismos que subyacen a gran parte de este tipo de fenómenos son desconocidos, por la enorme cantidad de variables incontrolables que en ellos intervienen, el uso del modelo normal puede justificarse asumiendo que cada observación se obtiene como la suma de unas pocas causas independientes.

De hecho, la estadística descriptiva sólo permite describir un fenómeno, sin explicación alguna. Para la explicación causal es preciso el diseño experimental, de ahí que al uso de la estadística en psicología y sociología sea conocido como método correlacional.

La distribución normal también es importante por su relación con la estimación por mínimos cuadrados, uno de los métodos de estimación más simples y antiguos.

Algunos ejemplos de variables asociadas a fenómenos naturales que siguen el modelo de la normal son:

- caracteres morfológicos de individuos como la estatura;
- caracteres fisiológicos como el efecto de un fármaco;
- caracteres sociológicos como el consumo de cierto producto por un mismo grupo de individuos;
- caracteres psicológicos como el cociente intelectual;
- nivel de ruido en telecomunicaciones;
- errores cometidos al medir ciertas magnitudes;
- etc.

La distribución normal también aparece en muchas áreas de la propia estadística. Por ejemplo, la distribución muestral de las medias muestrales es aproximadamente normal, cuando la distribución de la población de la cual se extrae la muestra no es normal.4 Además, la distribución normal maximiza la entropía entre todas las distribuciones con media y varianza conocidas, lo cual la convierte en la elección natural de la distribución subyacente a una lista de datos resumidos en términos de media muestral y varianza. La distribución normal es la más extendida en estadística y muchos tests estadísticos están basados en una "normalidad" más o menos justificada de la variable aleatoria bajo estudio.

En probabilidad, la distribución normal aparece como el límite de varias distribuciones de probabilidad continuas y discretas.

La distribución normal fue presentada por primera vez por Abraham de Moivre en un artículo del año 1733, que fue reimpreso en la segunda edición de su The Doctrine of Chances, de 1738, en el contexto de cierta aproximación de la distribución binomial para grandes valores de n. Su resultado fue ampliado por Laplace en su libro Teoría analítica de las probabilidades (1812), y en la actualidad se llama Teorema de De Moivre-Laplace.

Laplace usó la distribución normal en el análisis de errores de experimentos. El importante método de mínimos cuadrados fue introducido por Legendre en 1805. Gauss, que afirmaba haber usado el método desde 1794, lo justificó rigurosamente en 1809 asumiendo una distribución normal de los errores. El nombre de Gauss se ha asociado a esta distribución porque la usó con profusión cuando analizaba datos astronómicos6 y algunos autores le atribuyen un descubrimiento independiente del de De Moivre. Esta atribución del nombre de la distribución a una persona distinta de su primer descubridor es un claro ejemplo de la ley de Stigler.

El nombre de "campana" viene de Esprit Jouffret que usó el término "bell surface" (superficie campana) por primera vez en 1872 para una distribución normal bivariante de componentes independientes. El nombre de "distribución normal" fue otorgado independientemente por Charles S. Peirce, Francis Galton y Wilhelm Lexis hacia 1875. [cita requerida] A pesar de esta terminología, otras distribuciones de probabilidad podrían ser más apropiadas en determinados contextos.

La distribución normal es un ejemplo importante referido a una variable aleatoria continua (la variable puede tomar cualquier valor real).

Podemos usar la distribución normal como una herramienta para calcular probabilidades. Por ejemplo, puede usarse para aproximar la distribución binomial (calcular probabilidades de la distribución binomial con números 'grandes' no ha sido tarea sencilla). Esta propiedad está en el origen de la curva normal.

2.4.4. Otras distribuciones continúas.

En teoría de la probabilidad una distribución de probabilidad se llama continua si su función de distribución es continua. Puesto que la función de distribución de una variable aleatoria X viene dada por la definición implica que en una distribución de probabilidad continua X se cumple P[X = a] = 0 para todo número real a, esto es, la probabilidad de que X tome el valor a es cero para cualquier valor de a. Si la distribución de X es continua, se llama a X variable aleatoria continua.

Para una variable continua hay infinitos valores posibles de la variable y entre cada dos de ellos se pueden definir infinitos valores más. En estas condiciones no es posible deducir la probabilidad de un valor puntual de la variable; como se puede hacer en el caso de variables discretas, pero es posible calcular la probabilidad acumulada hasta un cierto valor (función de distribución de probabilidad), y se puede analizar cómo cambia la probabilidad acumulada en cada punto (estos cambios no son probabilidades sino otro concepto: la función de densidad.

Las distribuciones de probabilidad son idealizaciones de los polígonos de frecuencias. En el caso de una variable estadística continua consideramos el histograma de frecuencias relativas, y se comprueba que al aumentar el número de datos y el número de clases el histograma tiende a estabilizarse llegando a convertirse su perfil en la gráfica de una función.

Las distribuciones de probabilidad de variable continua se definen mediante una función y=f(x) llamada función de probabilidad o función de densidad.

Así como en el histograma la frecuencia viene dada por el área, en la función de densidad la probabilidad viene dada por el área bajo la curva, por lo que:

- El área encerrada bajo la totalidad de la curva es 1.
- Para obtener la probabilidad p(a£X£b) obtenemos la proporción de área que hay bajo la curva desde a hasta b.
- La probabilidad de sucesos puntuales es 0, p(X=a)=0.

FUNCIÓN DE DENSIDAD Y FUNCIÓN DE DISTRIBUCIÓN

Llamaremos función de densidad de una variable aleatoria continua X a una función f que cumple:

- Es positiva.
- El área total bajo la curva, es decir entre f(x) y el eje de abscisas, es 1.
- El área determinada por f(x), el eje de abscisas y las rectas x=a, x=b, es la probabilidad de que la variable continua X esté en el intervalo [a,b], P(a£X£b).

PARÁMETROS DE UNA DISTRIBUCIÓN DE PROBABILIDAD

Por analogía con las variables estadísticas podemos definir también aquí la media m y la desviación típica s de la variable aleatoria.

La media m, también llamada esperanza matemática, es un valor representativo de todos los valores que toma la variable aleatoria X, lo podemos imaginar como el punto sobre el eje de abscisas donde al poner una cuña la figura plana definida por la función de densidad quedará en equilibrio.

La desviación típica s es una medida de la dispersión de los valores que toma la variable aleatoria respecto de la media. Como ocurría con las variables estadísticas la desviación típica será más pequeña o más grande según la gráfica de la función de densidad sea más estrecha o más ancha en torno a la media.

La distribución Gamma

Este modelo es una generalización del modelo Exponencial ya que, en ocasiones, se utiliza para modelar variables que describen el tiempo hasta que se produce p veces un determinado suceso.

Propiedades de la distribución Gamma

- Su esperanza es pα.
- Su varianza es pα2.
- La distribución Gamma (α, p = 1) es una distribución Exponencial de parámetro α. Es decir, el modelo Exponencial es un caso particular de la Gamma con p = 1.
- Dadas dos variables aleatorias con distribución Gamma y parámetro α común.

Una consecuencia inmediata de esta propiedad es que, si tenemos k variables aleatorias con distribución Exponencial de parámetro α (común) e independientes, la suma de todas ellas seguirá una distribución G(α, k).

La distribución de Cauchy

Se trata de un modelo continuo Cuya integral nos proporciona la función de distribución.

Propiedades de la distribución de Cauchy

Se trata de un ejemplo de variable aleatoria que carece de esperanza (y, por tanto, también de varianza o cualquier otro momento).

Por tanto, la esperanza de una distribución de Cauchy no existe. Cabe señalar que la función de densidad es simétrica respecto al valor cero (que sería la mediana y la moda), pero al no existir la integral anterior, la esperanza no existe.

La distribución de Weibull

Se trata de un modelo continúo asociado a variables del tipo tiempo de vida, tiempo hasta que un mecanismo falla, etc.

Propiedades de la distribución Weibull

- Si tomamos $\beta = 1$ tenemos una distribución Exponencial.

UNIDAD III

3.1. Muestreo aleatorio simple.

El muestreo aleatorio simple (M.A.S.) es la técnica de muestreo en la que todos los elementos que forman el universo y que por lo tanto están descritos en el marco muestral, tienen idéntica probabilidad de ser seleccionados para la muestra. Sería algo así como hacer un sorteo justo entre los individuos del universo: asignamos a cada persona un boleto con un número correlativo, introducimos los números en una urna y empezamos a extraer al azar boletos. Todos los individuos que tengan un número extraído de la urna formarían la muestra. Obviamente, en la práctica, estos métodos pueden automatizarse mediante el uso de ordenadores.

Dependiendo de si los individuos del universo pueden ser seleccionados más de una vez en la muestra, hablaremos de M.A.S. con reposición o sin reposición.

Si usamos reposición, el hecho de que seleccione un individuo al azar para la muestra no impediría que este mismo individuo pudiese volver a ser seleccionado en una siguiente selección. Sería equivalente a decir que cada vez que se extrae un número al azar de la urna, volvemos a colocar el número antes de la siguiente extracción. Si por el contrario no usamos reposición, un individuo seleccionado para la muestra una vez ya no entraría nuevamente en el sorteo.

César Pérez López, en su libro "Muestreo Estadístico" (Pearson, 2005) desarrolla de forma muy clara una comparación entre ambas técnicas. Tanto si lo miramos desde el punto de vista de qué técnica genera estimaciones más precisas como desde el punto de vista de qué técnica permite tener la misma precisión con menor tamaño de muestra, se puede concluir que el muestreo aleatorio simple sin reposición siempre es más eficiente.

Para poder observar este resultado, partimos de la siguiente expresión para el tamaño de muestra en un M.A.S. sin reposición. La fórmula relaciona el tamaño de muestra necesario cuando el universo es finito con el tamaño necesario cuando el universo es infinito.

Beneficios del muestreo aleatorio simple

El desarrollo de la informática ha permitido que diseñar una muestra aleatoria simple sea extremadamente rápido y fiable. La generación de números aleatorios mediante software (estrictamente son números pseudo-aleatorios) es cada vez más fiable.

De esta forma, al usar M.A.S. nos aseguramos la obtención de muestras representativas, de manera que la única fuente de error que va a afectar a mis resultados va a ser el azar. Y lo que es más importante, este error debido al azar puede calcularse de forma precisa (o al menos acotarse).

Inconvenientes del muestreo aleatorio simple

El único inconveniente del M.A.S. es la dificultad de llevarlo a la práctica en investigaciones reales. Recordemos: al ser una técnica probabilística, es necesario un marco muestral con todos los individuos y que todos ellos sean seleccionables para la muestra. Un requisito que difícilmente puede cumplirse en la mayoría de estudios de mercado y opinión reales, lo que nos obligará a emplear otras técnicas.

El muestreo aleatorio simple es un procedimiento de muestreo probabilístico que da a cada elemento de la población objetivo y a cada posible muestra de un tamaño determinado, la misma probabilidad de ser seleccionado.

El muestreo aleatorio simple no es tan utilizado en investigaciones del consumidor, sobre todo porque es complicado obtener un marco de muestreo donde extraer al azar y no querrás darle a todas las unidades de la muestra una probabilidad igual de ser elegidas, ya que usualmente para hacer una investigación de este tipo se requiere a usuarios de tiendas o consumidores de ciertos productos o ciertas áreas específicas para ser las unidades de muestreo.

No olvidemos que una parte muy importante del muestreo consiste en tener el tamaño de la muestra correcta, para no tener un error de muestreo, el cual debe ser el mínimo posible.

Pasos para seleccionar una muestra aleatoria simple

- Define la población objetivo. Quizá quieras leer: ¿Cómo encontrar a tu mercado objetivo?
- Identifica un marco de muestreo actual de la población objetivo o desarrolla uno nuevo.
- Evalúa el marco de muestreo para la falta de cobertura, cobertura excesiva, cobertura múltiple y la agrupación, y haz los ajustes que consideres necesarios.
- Asigna un número único a cada elemento de la trama.
- Determina el tamaño de la muestra.
- Selecciona al azar el número específico de elementos de la población.

Para seleccionar el número de elementos de la población puedes recurrir al método de lotería, una tabla de números aleatorios y los números generados de forma aleatoria mediante un programa de computadora, es decir, al azar.

El método de lotería sólo funciona bien con pequeñas poblaciones de la muestra, es poco práctico para su uso con poblaciones más grandes.

El uso de números aleatorios, un método alternativo implica también la numeración de miembros de la población de 1 a N. Luego, el tamaño de muestra de n tiene que ser determinada por selección de los números al azar.

Los números que el investigador encuentra que no concuerdan con los números asignados a elementos de la población objetivo son ignorados. Este proceso de la tabla de números aleatorios es un proceso tedioso, consume tiempo, y no se recomienda para grandes poblaciones.

En su lugar, se pueden utilizar softwares estadísticos u hojas de cálculo para generar números aleatorios. Los elementos de las poblaciones cuyos números asignados coinciden con los números generados por el software son incluido en la muestra. Se puede seleccionar un número de una tabla de números aleatorios para usarlo como el número de partida para el procedimiento.

Subtipos de muestreo aleatorio simple

Hay dos tipos de muestreo aleatorio simple: el muestreo con reemplazo y sin reemplazo. En el muestreo con reemplazo, después de que un elemento ha sido seleccionado de entre el marco de la muestra se devuelve y es elegible para ser seleccionado de nuevo.

En el muestreo sin reemplazo, después de que un elemento se selecciona del marco de la muestra, se retira de la población y no regresa a la base del muestreo. Este tipo de muestreo suele ser más eficiente pues no permite que el mismo elemento de la población entre a la muestra más de una vez.

Ventajas y desventajas del muestreo aleatorio simple

- Entre sus puntos fuertes están que tiende a producir muestras representativas y permite el uso de la estadística inferencial en el analisis de datos recogidos.

- Cada selección es independiente de otras selecciones; Todas las combinaciones posibles de unidades de muestreo tienen la misma oportunidad de ser seleccionadas. En el muestreo sistemático, las posibilidades de ser seleccionado no son independientes entre sí.
- En general, es más fácil que otros procedimientos de muestreo probabilístico (tales como el muestreo por conglomerados) de comprender y comunicar a otros.
- Los procedimientos estadísticos requeridos para analizar los datos y calcular los errores son más fáciles que los requeridos en otros procedimientos de muestreo probabilístico. Te recomiendo leer: Tipos de errores. Consejos para que tu investigación no caiga en ellos
- Entre las desventajas están que se requiere un marco de muestreo de elementos de la población objetivo. Un marco de muestreo apropiado puede que no exista para la población que se dirige, y puede que no sea factible o práctico construir uno. En este caso el muestreo por conglomerados no requiere de una toma de muestra de los elementos de la población objetivo.
- El muestreo aleatorio simple tiende a tener errores de muestreo más grandes y menos precisión de muestreo estratificado del mismo tamaño de la muestra.
- Los encuestados pueden estar muy dispersos, por tanto, los costos de la recolección de datos pueden ser superiores a las de otros diseños de la muestra de probabilidad, como el muestreo por conglomerados.
- El muestreo aleatorio simple puede no producir un número suficiente de elementos de pequeños subgrupos. Esto no haría de un muestreo aleatorio simple una buena opción para los estudios que requieren un

análisis comparativo de las categorías pequeñas de una población con categorías mucho más amplias de la población.

3.1.1. Justificación del muestreo.

En vez de tomar un censo completo, los procedimientos de muestreo estadístico se han convertido en la herramienta preferida en la mayoría de las situaciones de investigación. Existen tres razones principales para extraer una muestra.

Antes que todo, por lo general, lleva demasiado tiempo realizar un censo completo. En segundo lugar, es demasiado costoso hacer un censo completo. Tercero, es demasiado molesto e ineficiente obtener un conteo completo de la población objeto.

Después que se han determinado las preguntas numéricas y categóricas más esenciales en la encuesta, el tamaño de muestra necesario se basará en la satisfacción de la pregunta con los requerimientos más rigurosos.

3.1.2. Función de Distribución empírica.

Los tratamientos estadísticos se carácterizan por un ir y venir permanente entre los datos, que son colecciones de cifras medidas, y los modelos probabilistas que no tienen ninguna realidad física, pero proveen herramientas para describir la variabilidad de los datos. En esta manera de pensar, un primer paso consiste en asociar a la muestra una ley de probabilidad ficticia. La distribución empírica asociada a una muestra es la ley de probabilidad sobre el conjunto de las modalidades, que afecta a cada observación con el peso 1/n.

La media, la varianza y la desviación estándar pueden ser vistas como características probabilistas de la distribución empírica. La media de la muestra es la esperanza de su distribución empírica.

Para un carácter discreto, la moda de la distribución empírica es el valor que tiene la frecuencia más alta. Para un carácter continúo agrupado en clases de amplitudes iguales, hablamos de clase modal. Una distribución empírica se llama unimodal si la frecuencia maximal es significativamente mayor que las otras. Puede ser bimodal o multimodal en otros casos.

Para estudiar una distribución empírica, la primera etapa consiste en ordenar los datos en orden creciente, es decir escribir sus estadígrafos de orden.

La función de distribución empírica (FED) o cdf empírica es una función de paso que salta por 1/N a la ocurrencia de cada observación.

Por definición, la función FDE calcula la distribución acumulativa del número aleatorio subyacente.

El FED estima la verdadera función de densidad acumulativa subyacente de los puntos en la muestra; Se garantiza virtualmente que converge con la distribución verdadera a medida que el tamaño de la muestra se hace lo suficientemente grande.

Si se tiene una muestra aleatoria simple, es posible encontrar una distribución a partir de la muestra que proporcionará un cierto parecido a la distribución verdadera de la variable asociada con la población.

Es lo que se denomina función de distribución empírica de la muestra.

La principal propiedad de la función de distribución empírica de la muestra es su aproximación a la función de distribución poblacional cuando aumenta el tamaño muestral.

Ello es conocido en estadística como el teorema de Glivenko-Cantelli o también como teorema central de la estadística.

3.1.3. Estadísticos muéstrales. Distribuciones.

En estadística un estadístico (muestral) es una medida cuantitativa, derivada de un conjunto de datos de una muestra, con el objetivo de estimar o inferir características de una población o modelo estadístico.

Más formalmente un estadístico es una función medible T que, dada una muestra estadística de valores, les asigna un número, que sirve para estimar determinado parámetro de la distribución de la que procede la muestra. Así, por ejemplo, la media de los valores de una muestra (media muestral) sirve para estimar la media de la población de la que se ha extraído la misma; la varianza muestral podría usarse para estimar la varianza poblacional, etc. Esto se denomina como realizar una estimación puntual.

A partir de las muestras seleccionadas de una población pueden construirse variables aleatorias alternativas, de cuyo análisis se desprenden interesantes propiedades estadísticas. Las dos formas más comunes de estas variables corresponden a las distribuciones muéstrales de las medias y de las proporciones.

Dada una población constituida por un número n de elementos, cuya media aritmética es m y donde la desviación típica viene dada s, pueden formarse n2 muestras con reemplazamiento distintas, formadas por dos elementos de la población.

El estudio de determinadas características de una población se efectúa a través de diversas muestras que pueden extraerse de ella.

El muestreo puede hacerse con o sin reposición, y la población de partida puede ser infinita o finita. Una población finita en la que se efectúa muestreo con reposición puede considerarse infinita teóricamente. También, a efectos prácticos, una población muy grande puede considerarse como infinita. En todo nuestro estudio vamos a limitarnos a una población de partida infinita o a muestreo con reposición.

Consideremos todas las posibles muestras de tamaño n en una población. Para cada muestra podemos calcular un estadístico (media, desviación típica, proporción,...) que variará de una a otra. Así obtenemos una distribución del estadístico que se llama distribución muestral.

Las dos medidas fundamentales de esta distribución son la media y la desviación típica, también denominada error típico.

Hay que hacer notar que si el tamaño de la muestra es lo suficientemente grande las distribuciones muéstrales son normales y en esto se basarán todos los resultados que alcancemos.

3.2. Estimación.

La estimación puntual consiste en atribuir un valor (la estimación) al parámetro poblacional. Si la muestra es representativa de la población, podemos esperar que los estadísticos calculados en las muestras tengan valores semejantes a los parámetros poblacionales, y la estimación consiste en asignar los valores de los estadísticos muéstrales a los parámetros poblacionales. Los estadísticos con que obtenemos las estimaciones se denominan estimadores.

En inferencia estadística se llama estimación al conjunto de técnicas que permiten dar un valor aproximado de un parámetro de una población a partir de los datos proporcionados por una muestra. Por ejemplo, una estimación de la media de una determinada característica de una población de tamaño N podría ser la media de esa misma característica para una muestra de tamaño n.

La estimación se divide en tres grandes bloques, cada uno de los cuales tiene distintos métodos que se usan en función de las características y propósitos del estudio:

Estimación puntual:

Método de los momentos;

Método de la máxima verosimilitud;

Método de los mínimos cuadrados;

Estimación por intervalos.

Estimación bayesiana.

Un estimador es una regla que establece cómo calcular una estimación basada en las mediciones contenidas en una muestra estadística.

Estimación puntual

Consiste en la estimación del valor del parámetro mediante un sólo valor, obtenido de una fórmula determinada. Por ejemplo, si se pretende estimar la talla media de un determinado grupo de individuos, puede extraerse una muestra y ofrecer como estimación puntual la talla media de los individuos. Lo más importante de un estimador, es que sea un estimador eficiente. Es decir, que sea insesgado (ausencia de sesgos) y estable en el muestreo o eficiente (varianza mínima) Estimación puntual. Sea X una variable poblacional con distribución $F\theta$, siendo θ desconocido. El problema de estimación puntual consiste en, seleccionada una muestra $X1, ..., Xn$, encontrar el estadístico $T(X1, ..., Xn)$ que mejor estime el parámetro θ. Una vez observada o realizada la muestra, con valores $x1, ..., xn$, se obtiene la estimación puntual de θ, $T(x1, ..., xn) = \hat{}\ \theta$.

Estimación por intervalos

Consiste en la obtención de un intervalo dentro del cual estará el valor del parámetro estimado con una cierta probabilidad. En la estimación por intervalos se usan los siguientes conceptos:

Intervalo de confianza

El intervalo de confianza es una expresión del tipo $[\theta 1, \theta 2]$ ó $\theta 1 \leq \theta \leq \theta 2$, donde θ es el parámetro a estimar. Este intervalo contiene al parámetro estimado con un determinado nivel de confianza. Pero a veces puede cambiar este intervalo cuando la muestra no garantiza un axioma o un equivalente circunstancial.

Variabilidad del Parámetro

Si no se conoce, puede obtenerse una aproximación en los datos aportados por la literatura científica o en un estudio piloto. También hay métodos para calcular el tamaño de la muestra que prescinden de este aspecto. Habitualmente se usa como medida de esta variabilidad la desviación típica poblacional y se denota σ.

Error de la estimación

Es una medida de su precisión que se corresponde con la amplitud del intervalo de confianza. Cuanta más precisión se desee en la estimación de un parámetro, más estrecho deberá ser el intervalo de confianza y, si se quiere mantener o disminuir el error, más observaciones deberán incluirse en la muestra estudiada. En caso de no incluir nuevas observaciones para la muestra, más error se comete al aumentar la precisión. Se suele llamar E, según la fórmula $E = (\theta 2 - \theta 1)/2$.

Límite de Confianza

Es la probabilidad de que el verdadero valor del parámetro estimado en la población se sitúe en el intervalo de confianza obtenido. El nivel de confianza se denota por $(1-\alpha)$, aunque habitualmente suele expresarse con un porcentaje $((1-\alpha) \cdot 100\%)$. Es habitual tomar como nivel de confianza un 95% o un 99%, que se corresponden con valores α de 0,05 y 0,01 respectivamente.

Valor α

También llamado nivel de significación. Es la probabilidad (en tanto por uno) de fallar en nuestra estimación, esto es, la diferencia entre la certeza (1) y el nivel de confianza $(1-\alpha)$. Por ejemplo, en una estimación con un nivel de confianza del 95%, el valor α es (100-95)/100 = 0,05.

Valor crítico

Se representa por $Z\alpha/2$. Es el valor de la abscisa en una determinada distribución que deja a su derecha un área igual a $\alpha/2$, siendo $1-\alpha$ el nivel de confianza. Normalmente los valores críticos están tabulados o pueden calcularse en función de la distribución de la población. Por ejemplo, para una distribución normal, de media 0 y desviación típica 1, el valor crítico para α = 0,1 se calcularía del siguiente modo: se busca en la tabla de la distribución ese valor (o el más aproximado), bajo la columna "Área"; se observa que se corresponde con -1,28. Entonces $Z\alpha/2$ = 1,64. Si la media o desviación típica de la distribución normal no coinciden con las de la tabla, se puede realizar el cambio de variable $t = (X-\mu)/\sigma$ para su cálculo.

Otros usos del término

El término estimación también se utiliza en ciencias aplicadas para hacer referencia a un cálculo aproximado, que normalmente se apoya en la herramienta estadística, aunque puede no hacerlo. En este sentido, un ejemplo clásico son los poco conocidos pero útiles en economía problemas de Fermi.

3.2.1. Propiedades de los estimadores.

ESTIMADOR:

Es un estadístico (es decir, es una función de la muestra) usado para estimar un parámetro desconocido de la población. Por ejemplo, si se desea conocer el precio medio de un artículo (el parámetro desconocido) se recogerán observaciones del precio de dicho artículo en diversos establecimientos (la muestra) y la media aritmética de las observaciones puede utilizarse como estimador del precio medio.

Para cada parámetro pueden existir varios estimadores diferentes. En general, escogeremos el estimador que posea mejores propiedades que los restantes, como insesgadez, eficiencia, convergencia y robustez (consistencia).

SESGO:

Se denomina sesgo de un estimador a la diferencia entre la esperanza (o valor esperado) del estimador y el verdadero valor del parámetro a estimar. Es deseable que un estimador sea insesgado o centrado, es decir, que su sesgo sea nulo por ser su esperanza igual al parámetro que se desea estimar.

Por ejemplo, si se desea estimar la media de una población, la media aritmética de la muestra es un estimador insesgado de la misma, ya que su esperanza (valor esperado) es igual a la media de la población.

EFICIENCIA:

Un estimador es más eficiente o preciso que otro, si la varianza del primero es menor que la del segundo.

CONVERGENCIA:

Para estudiar las características de un estimador no solo basta con saber el sesgo y la varianza, sino que además es útil hacer un análisis de su comportamiento y estabilidad en el largo plazo, esto es, su comportamiento asintótico. Cuando hablamos de estabilidad en largo plazo, se viene a la mente el concepto de convergencia. Luego, podemos construir sucesiones de estimadores y estudiar el fenómeno de la convergencia.

Comportamiento Asintótico: En el caso de las variables aleatorias, existen diversos tipos de convergencia, dentro de las cuales podemos distinguir:

-Convergencia en probabilidad (o débil).

-Convergencia casi segura (o fuerte).

-Convergencia en media cuadrática.

-Convergencia en distribución.

CONSISTENCIA:

También llamada robustez, se utilizan cuando no es posible emplear estimadores de mínima varianza, el requisito mínimo deseable para un estimador es que a medida que el tamaño de la muestra crece, el valor del estimador tiende a ser el valor del parámetro, propiedad que se denomina consistencia.

Las propiedades de los estimadores son las cualidades que pueden tener estos y que sirven para escoger aquellos que son más capaces de arrojar buenos resultados.

Por empezar definiendo el concepto de estimador, diremos que dada una muestra aleatoria cualquiera {x1,x2,x3, … ,xn} un estimador representa a una población que depende de ϕ un parámetro que desconocemos.

Este parámetro al que nosotros denotamos con la letra griega fi (ϕ), puede ser, por ejemplo, la media de una variable aleatoria cualquiera.

Matemáticamente, un estimador Q de un parámetro depende de las observaciones aleatorias de la muestra {x1,x2,x3, … ,xn} y de una función conocida (h) de la muestra. El estimador (Q) será una variable aleatoria porque depende de la muestra que contiene variables aleatorias.

insesgadez de un estimador

Un estimador Q de ϕ es un estimador insesgado si $E(Q)= \phi$ para todos los valores posibles de ϕ . Definimos E(Q) como el valor esperado o esperanza del estimador Q.

En el caso de presencia de estimadores sesgados, este sesgo se representaría como:

$Sesgo (Q) = E(Q) – \phi$

Podemos ver que el sesgo es la diferencia entre el valor esperado del estimador, E(Q), y el verdadero valor del parámetro poblacional, ϕ.

Eficiencia de un estimador

Si Q1 y Q2 son dos estimadores insesgados de ϕ , será eficiente su relación con Q2 cuando $Var(Q1) \leq Var(Q2)$ para cualquier valor de ϕ siempre que la muestra estadística de ϕ sea estrictamente mayor a 1, $n>1$. Siendo Var, la varianza y n, el tamaño de la muestra.

Dicho de forma intuitiva, suponiendo que tenemos dos estimadores con la propiedad de insesgadez, podemos decir que uno (Q1) es más eficiente que otro (Q2) si la variabilidad de los resultados de uno (Q1) es menor que la del otro (Q2). Es lógico pensar que una cosa que varía más que otra es menos ‘precisa’.

Por tanto, sólo podemos usar este criterio de selección de estimadores cuando son insesgados. En el enunciado anterior cuando estamos definiendo la eficiencia ya suponemos que los estimadores tienen que ser insesgados.

Para comparar estimadores que no son necesariamente insesgados, esto es, que puede existir sesgo, se recomienda calcular el Error Cuadrático Medio (ECM) de los estimadores.

El Error Cuadrático Medio (ECM) calcula la distancia promedio que existe entre el valor esperado del estimador muestral Q y el estimador poblacional. La forma cuadrática del ECM se debe a que los errores pueden ser por defecto, negativos, o por exceso, positivos, respecto al valor esperado. De este modo, ECM siempre computará valores positivos.

ECM depende de la varianza y del sesgo (en el caso que lo hubiera) permitiéndonos comparar dos estimadores cuando uno o ambos son sesgados. Aquel cuyo ECM sea mayor se entenderá que es menos preciso (tiene más error) y, por tanto, menos eficiente.

Consistencia de un estimador

La consistencia es una propiedad asintótica. Esta propiedad se parece a la propiedad de eficiencia con la diferencia de que la consistencia mide la distancia probable entre el valor del estimador y el valor verdadero del parámetro poblacional a medida que el tamaño de la muestra aumenta indefinidamente. Este aumento indefinido del tamaño de la muestra es la base de la propiedad asintótica.

Existe una dimensión de muestra mínima para llevar a cabo el análisis asintótico (comprobar la consistencia del estimador a medida que aumenta la muestra).

Aproximaciones de muestras grandes funcionan correctamente para muestras de alrededor de 20 observaciones, (n = 20). En otras palabras, queremos ver cómo se comporta el estimador cuando aumentamos la muestra, pero este aumento tiende a infinito. Dado esto, hacemos una aproximación y a partir de 20 observaciones en una muestra (n ≥ 20), es apropiado el análisis asintótico.

Esto nos dice que las diferencias entre el estimador y su valor poblacional, $|Qn - \phi|$, tienen que ser mayores que cero. Por esto lo expresamos en valor absoluto. La probabilidad de esta diferencia tiende a 0 (se hace cada vez más pequeña) cuando el tamaño de la muestra (n) tiende a infinito (se hace cada vez más grande).

3.2.2. Obtención de estimadores.

Método de los momentos

Se trata de un método de obtención de estimadores muy intuitivo. Básicamente, consiste en igualar los momentos poblacionales (que sean función del o los parámetros a estimar) con los momentos muéstrales y despejar el parámetro a estimar.

Así, por ejemplo, la esperanza de una variable aleatoria se estimaría por la media muestral; la varianza, por la varianza muestral; etc.

La principal ventaja de este método es su simplicidad. Sin embargo, aunque los estimadores así obtenidos son consistentes, en general, no son centrados ni eficientes. Además, en ciertos casos puede proporcionar estimaciones absurdas.

Supongamos que tenemos una variable con distribución uniforme donde el límite inferior es cero y el superior es desconocido. Naturalmente, estaremos interesados en estimar el límite superior (al que llamaremos b) de nuestra distribución uniforme.

Recordemos que la esperanza de una distribución uniforme comprendida entre dos valores a y b es el promedio de estos dos valores.

Por tanto, para aplicar el método de los momentos para estimar b, igualaremos dicho promedio a la media aritmética.

Ventajas y desventajas del método de los momentos

El método de los momentos es bastante sencillo y brinda estimadores compatibles (debajo suposiciones muy débiles), aunque estos estimadores son a menudo sesgados.

En algunos casos, cuándo estimamos parámetros de una familia conocida de distribuciones de probabilidad, este método es sustituido el método de máxima verosimilitud de Fisher , porque con máxima verosimilitud los estimadores tienen probabilidad más alta de ser cercanos a las cantidades que estimamos y son menos sesgadas.

Aun así, en algunos casos las ecuaciones del método de máxima verosimilitud pueden ser intratables sin ayuda de ordenadores, mientras que el método de estimadores de los momentos puede ser más accesible y fácilmente calculado a mano.

Las estimaciones por el método de los momentos pueden ser utilizadas como la primera aproximación a las soluciones de las ecuaciones de verosimilitud, y podemos encontrar sucesivas mejoras en las aproximaciones por el método de

Newton-Raphson. De este modo el método de momentos puede ayudar a encontrar estimaciones del método de máxima verosimilitud.

En algunos casos, infrecuentes con muestras grandes, pero no tan infrecuentes con muestras pequeñas, las estimaciones dadas por el método de momentos están por fuera del espacio paramétrico, por lo que no tiene sentido confiar en ellos. Este problema nunca surge en el método de máxima verosimilitud. También, estimaciones del método de los momentos no son necesariamente estadísticos suficientes, p.e., a veces fallan en tener en cuenta toda información pertinente en la muestra.

Cuándo estimamos otros parámetros estructurales (p. ej., parámetros de una función de utilidad, en vez de parámetros de una distribución de probabilidad sabida), las distribuciones de probabilidad apropiadas pueden ser desconocidas, por lo que en tal caso es preferible el método de los momentos al de máxima verosimilitud.

3.2.3. Estimación por intervalos de confianza.

La estadística inferencial es el proceso de uso de los resultados derivados de las muestras para obtener conclusiones acerca de las características de una población. La estadística inferencial nos permite estimar características desconocidas como la media de la población o la proporción de la población. Existen dos tipos de estimaciones usadas para estimar los parámetros de la población: la estimación puntual y la estimación de intervalo. Una estimación puntual es el valor de un solo estadístico de muestra. Una estimación del intervalo de confianza es un rango de números, llamado intervalo, construido alrededor de la estimación puntual. El intervalo de confianza se construye de manera que la probabilidad del parámetro de la población se localice en algún lugar dentro del intervalo conocido.

La estimación por intervalos consiste en establecer el intervalo de valores donde es más probable se encuentre el parámetro. La obtención del intervalo se basa en las siguientes consideraciones:

a) Si conocemos la distribución muestral del estimador podemos obtener las probabilidades de ocurrencia de los estadísticos muéstrales.

b) Si conociéramos el valor del parámetro poblacional, podríamos establecer la probabilidad de que el estimador se halle dentro de los intervalos de la distribución muestral.

c) El problema es que el parámetro poblacional es desconocido, y por ello el intervalo se establece alrededor del estimador. Si repetimos el muestreo un gran número de veces y definimos un intervalo alrededor de cada valor del estadístico muestral, el parámetro se sitúa dentro de cada intervalo en un porcentaje conocido de ocasiones. Este intervalo es denominado "intervalo de confianza".

La estimación puntual trata el problema de estimar mediante un número el valor de una característica poblacional o parámetro θ desconocido (por ejemplo, la estimación del IPC de un determinado periodo).

En muchos casos la estimación puntual no es suficiente en el sentido de que no nos indica el error que se comete en dicha estimación.

Lo razonable en la práctica es adjuntar, junto a la estimación puntual del parámetro, un cierto intervalo numérico que mida el margen de error que, de acuerdo a las observaciones muéstrales, pueda tener dicha estimación.

La idea de Intervalo de Confianza, es proponer un rango de valores entre los que posiblemente se encuentre el verdadero valor del parámetro θ.

3.3. Contraste de hipótesis.

Dentro de la inferencia estadística, un contraste de hipótesis (también denominado test de hipótesis o prueba de significación) es un procedimiento para juzgar si una propiedad que se supone en una población estadística es compatible con lo observado en una muestra de dicha población. Fue iniciada por Ronald Fisher y fundamentada posteriormente por Jerzy Neyman y Karl Pearson.

Mediante esta teoría, se aborda el problema estadístico considerando una hipótesis determinada, y una hipótesis alternativa y se intenta dirimir cuál de las dos es la hipótesis verdadera, tras aplicar el problema estadístico a un cierto número de experimentos.

Está fuertemente asociada al concepto estadístico de potencia y a los conceptos de errores de tipo I y II, que definen respectivamente, la posibilidad de tomar un suceso falso como verdadero, o uno verdadero como falso.

Los tipos más importantes son los test centrados, de hipótesis y alternativa simple, aleatorizados, etc. Dentro de los tests no paramétricos, el más extendido es probablemente el test de la U de Mann-Whitney.

La aplicación de cálculos probabilísticos permite determinar a partir de qué valor debemos rechazar la hipótesis garantizando que la probabilidad de cometer un error es un valor conocido a priori. Las hipótesis pueden clasificarse en dos grupos, según:

- Especifiquen un valor concreto o un intervalo para los parámetros del modelo.
- Determinen el tipo de distribución de probabilidad que ha generado los datos.

Un ejemplo del primer grupo es la hipótesis de que la media de una variable es 10, y del segundo que la distribución de probabilidad es la distribución normal.

Aunque la metodología para realizar el contraste de hipótesis es análoga en ambos casos, distinguir ambos tipos de hipótesis es importante puesto que muchos problemas de contraste de hipótesis respecto a un parámetro son, en realidad, problemas de estimación, que tienen una respuesta complementaria dando un intervalo de confianza (o conjunto de intervalos de confianza) para dicho parámetro. Sin embargo, las hipótesis respecto a la forma de la distribución se suelen utilizar para validar un modelo estadístico para un fenómeno aleatorio que se está estudiando.

Se denomina hipótesis nula a la hipótesis que se desea contrastar. El nombre de "nula" significa "sin valor, efecto o consecuencia", lo cual sugiere que, debe identificarse con la hipótesis de no cambio (a partir de la opinión actual); no diferencia, no mejora, etc. Representa la hipótesis que mantendremos a no ser que los datos indiquen su falsedad, y puede entenderse, por tanto, en el sentido de "neutra". La hipótesis nunca se considera probada, aunque puede ser rechazada por los datos. Por ejemplo, la hipótesis de que dos poblaciones tienen la misma media puede ser rechazada fácilmente cuando ambas difieren mucho, analizando muestras suficientemente grandes de ambas poblaciones, pero no puede ser "demostrada" mediante muestreo, puesto que siempre cabe la posibilidad de que las medias difieran en una cantidad lo suficientemente pequeña para que no pueda ser detectada, aunque la muestra sea muy grande.

A partir de una muestra de la población en estudio, se extrae un estadístico (esto es, un valor que es función de la muestra) cuya distribución de probabilidad esté relacionada con la hipótesis en estudio y sea conocida. Se toma entonces como región de rechazo al conjunto de valores que es más improbable bajo la hipótesis, esto es, el conjunto de valores para el que rechazaremos la hipótesis nula si el valor del estadístico observado entra dentro de él.

La probabilidad de que se obtenga un valor del estadístico que entre en la región de rechazo aun siendo cierta la hipótesis puede calcularse. De esta manera, se puede escoger dicha región de tal forma que la probabilidad de cometer este error sea suficientemente pequeña.

POSIBLES ERRORES EN EL CONTRASTE DE HIPÓTESIS

El contraste de hipótesis no establece la verdad de la hipótesis, sino un criterio que nos permite decidir si una hipótesis se acepta o se rechaza, o el determinar si las muestras observadas difieren significativamente de los resultados esperados. En este proceso podemos incurrir en dos tipos de errores según sea la situación real y la decisión que tomemos.

Si rechazamos una hipótesis cuando debiera ser aceptada, cometemos un error de tipo I, mientras que si la aceptamos debiendo ser rechazada diremos que hemos cometido un error de tipo II. Minimizar los errores no es una cuestión sencilla, un tipo suele ser más grave que otro y los intentos de disminuir uno suelen producir el aumento del otro. La única forma de disminuir ambos a la vez es aumentar el tamaño de la muestra.

3.3.1. Concepto y definiciones.

Una hipótesis estadística es una afirmación respecto a alguna característica de una población. Contrastar una hipótesis es comparar las predicciones con la realidad que observamos. Si dentro del margen de error que nos permitimos admitir, hay coincidencia, aceptaremos la hipótesis y en caso contrario la rechazaremos.

La hipótesis emitida se suele designar por Ho y se llama Hipótesis nula porque parte del supuesto que la diferencias entre el valor verdadero del parámetro y su valor hipotético es debida al azar, es decir no hay diferencia.

La hipótesis contraria se designa por H1 y se llama Hipótesis alternativa

Los contrastes pueden ser unilaterales o bilaterales (también llamados de una o dos colas) según establezcamos las hipótesis, si las definimos en términos de igual y distinto estamos ante una hipótesis unilateral, si suponemos una dirección (en términos de mayor o menor) estamos ante uno unilateral.

Se trata pues, de extraer conclusiones a partir de una muestra aleatoria y significativa, que permitan aceptar o rechazar una hipótesis previamente emitida, sobre el valor de un parámetro desconocido de la población. El método que seguiremos es el siguiente:

Enunciar la hipótesis

Elegir un nivel de significación a y construir la zona de aceptación, intervalo fuera del cual sólo se encuentran el a100% de los casos más raros. A la zona de rechazo la llamaremos región crítica, y su área es el nivel de significación.

Verificar la hipótesis extrayendo una muestra cuyo tamaño se ha decidido en el paso anterior y obteniendo de ella el correspondiente estadístico (media o proporción en nuestro caso).

Decidir. Si el valor calculado en la muestra cae dentro de la zona de aceptación se acepta la hipótesis y si no se rechaza.

Una hipótesis estadística es una asunción relativa a una o varias poblaciones, que puede ser cierta o no. Las hipótesis estadísticas se pueden contrastar con la información extraída de las muestras y tanto si se aceptan como si se rechazan se puede cometer un error.

La hipótesis formulada con intención de rechazarla se llama hipótesis nula y se representa por H0. Rechazar H0 implica aceptar una hipótesis alternativa (H1).

3.3.2. Construcción de Test de hipótesis.

Seis pasos básicos para configurar y realizar correctamente una prueba de hipótesis.

Especificar las hipótesis.

Elegir un nivel de significancia (también denominado alfa o α).

El gerente selecciona un nivel de significancia de 0.05, que es el nivel de significancia más utilizado.

Determinar la potencia y el tamaño de la muestra para la prueba.

Recolectar los datos.

Comparar el valor p de la prueba con el nivel de significancia.

Después de realizar la prueba de hipótesis, el gerente obtiene un valor p de 0.004. El valor p es menor que el nivel de significancia de 0.05.

Decidir si rechazar o no rechazar la hipótesis nula.

3.4. Contraste de hipótesis paramétricas.

El problema del contraste de hipótesis consiste básicamente en comprobar cotejar, decidir, en definitiva, sobre la veracidad de una hipótesis prefijada previamente como supuestamente cierta. En términos estadísticos, la o las hipótesis que formulamos lo serán lógicamente sobre la población. Bien afectando a algún parámetro de ésta, lo que da origen a los contrastes paramétricos o bien a otras características de la mismas que no lo sean estrictamente, lo que origina contrates "no" paramétricos. Si bien este capítulo está dedicado a los contrastes paramétricos, esta introducción puede considerarse común a ambos tipos de contrastes.

La solución estadística del problema de contrastación se basará en los datos muéstrales y la base estadística (probabilística) de la que arrancará el contraste va a ser la distribución de algún estadístico muestral

Supongamos que deseamos hacer un contraste acerca de un parámetro, de la población. Para llevarlo a cabo consideraremos la distribución de algún estadístico muestral que de alguna manera se corresponda y se relacione con el parámetro; designemos en general a este estadístico como T. Si con los datos muéstrales obtenemos un valor concreto para T tal que pertenezca a una determinada región del campo de variación de T optaremos por no rechazar la hipótesis y en caso contrario por rechazarla. Obviamente la clave del problema será delimitar la región del campo de variación de T que consideraremos como zona de aceptación de la hipótesis. Esto se resolverá por un criterio probabilístico partiendo de la distribución muestral de T.

Pasemos a definir los principales conceptos implicados en nuestro problema:

Región crítica. Será aquella región del campo de variación del estadístico tal que si contiene al valor evaluado del mismo con los datos muéstrales nos llevará a rechazar la hipótesis . La designaremos por R1

Región de aceptación. Es la región complementaria de la anterior. Si el valor evaluado del estadístico pertenece a ella No rechazamos la hipótesis. (Las hipótesis nunca se aceptan de forma definitiva, sólo se aceptan provisionalmente, es decir, no se rechazan, a la espera de una nueva información que eventualmente pueda llevarnos a rechazarla en el futuro). La designaremos por R0.

Evidentemente los conjuntos de puntos que forman ambas regiones son disjuntos.

Una hipótesis estadística (paramétrica) es una conjetura sobre el valor concreto que tiene en realidad. El establecer una hipótesis sobre un parámetro, supone dividir los posibles valores del parámetro en dos grupos disjuntos tales que unos son hipotéticamente ciertos (0) y los otros (1) no lo son. A la hipótesis que se desea contrastar se la denomina "hipótesis nula", siendo, por tanto, el valor o valores 0 que hipotéticamente consideramos reales, dicha hipótesis viene expresada como H0. Alternativamente y consecuentemente se establece la denominada "hipótesis alternativa " (H1) compuesta ésta por el valor o valores 1 que en consecuencia de la elección y de la complementariedad de los de la hipótesis nula, son los que, en principio, no consideramos cómo hipotéticamente reales.

El hecho de que las hipótesis, tanto la nula cómo la alternativa pueda recoger en sus planteamientos uno o varios valores, da lugar a hipótesis de carácter simple, si el número de valores plausibles e hipotéticos es de uno en ambas, o bien a hipótesis compuestas si dicho valor no es único en alguna de ellas.

Teniendo en cuenta lo dicho anteriormente, el problema de rechazar o aceptar una hipótesis puede plantearse como un problema de decisión, en el que evidentemente existe la posibilidad de fracasar o acertar en la elección o decisión a la hora de concluir que la hipótesis, bien nula bien alternativa, son rechazables o no, dado, claro está, que no conocemos la verdad.

3.4.1. Test para poblaciones normales.

Una prueba de 2 muestras se puede utilizar para comparar si las medias de dos grupos independientes son diferentes. Esta prueba se deriva bajo el supuesto de que ambas poblaciones están normalmente distribuidas y poseen varianzas iguales. Si bien el supuesto de normalidad no es crítico (Pearson, 1931; Barlett, 1935; Geary, 1947), el supuesto de varianzas iguales es crítico si los tamaños de las muestras son notablemente diferentes (Welch, 1937; Horsnell, 1953).

Algunos profesionales primero realizan una prueba preliminar para evaluar varianzas iguales antes de realizar el clásico procedimiento t de 2 muestras. Sin embargo, este enfoque presenta serias desventajas debido a que estas pruebas de varianzas están sujetas a supuestos y limitaciones importantes. Por ejemplo, numerosas pruebas de varianzas iguales, como la clásica prueba F, son sensibles a desviaciones con respecto a la normalidad. Otras pruebas que no se basan en el supuesto de normalidad, como la de Levene/Brown-Forsythe, tienen poca potencia para detectar una diferencia entre varianzas.

B.L. Welch desarrolló un método de aproximación para comparar las medias de dos poblaciones normales independientes cuando las varianzas no son necesariamente iguales (Welch, 1947).

Debido a que la prueba t modificada de Welch no se deriva bajo el supuesto de varianzas iguales, los usuarios pueden comparar las medias de dos poblaciones sin primero tener que determinar la existencia de varianzas iguales.

Comparación entre la prueba de 2 muestras clásica y la prueba t de Welch

Si los datos provienen de dos poblaciones normales con las mismas varianzas, la prueba de 2 muestras clásica tiene la misma o más potencia que la prueba de Welch. El supuesto de normalidad no es crítico para el procedimiento clásico (Pearson, 1931; Barlett, 1935; Geary, 1947); sin embargo, el supuesto de varianzas iguales es importante para garantizar resultados válidos. Más específicamente, el procedimiento clásico es sensible al supuesto de varianzas iguales cuando difieren los tamaños de las muestras independientemente de qué tan grandes sean las muestras (Welch, 1937; Horsnell, 1953). En la práctica, sin embargo, rara vez se cumple el supuesto de varianzas iguales, lo cual puede producir tasas de error Tipo I más elevadas. Por lo tanto, si la prueba t de 2 muestras clásica se utiliza cuando dos muestras tienen varianzas diferentes, aumenta la probabilidad de que la prueba produzca resultados incorrectos.

La prueba t de Welch es una alternativa viable a la prueba t clásica debido que no asume varianzas iguales y, por lo tanto, no es sensible a varianzas desiguales para todos los tamaños de muestra. Sin embargo, la prueba t de Welch se basa en aproximaciones y su rendimiento con tamaños de muestra pequeños pudiera ser cuestionable. Queríamos determinar si la prueba t de Welch o la prueba t de 2 muestras clásica es la prueba más fiable y práctica para utilizar en el Asistente.

3.4.2. Test para poblaciones binomiales y de Poisson.

Nos encontramos con un modelo derivado de un proceso experimental puro, en el que se plantean las siguientes circunstancias.

Se realiza un número n de pruebas (separadas o separables).

Cada prueba puede dar dos únicos resultados A y Ã

La probabilidad de obtener un resultado A es p y la de obtener un resultado Ã es q, con q= 1-p, en todas las pruebas. Esto implica que las pruebas se realizan exactamente en las mismas condiciones y son , por tanto ,independientes en sus resultados. Si se trata de extracciones, (muestreo), las extracciones deberán ser con devolución (reemplazamiento) , o bien población grande (M.A.S). A este respecto hagamos una consideración: si el proceso consiste en extraer individuos de una población y observar si poseen cierta característica: el parámetro n será el número de extracciones (tamaño muestral) y el parámetro p la proporción de individuos de la población que poseen la característica en cuestión. Se ha comentado que para que la probabilidad, de que en cada extracción obtengamos un individuo poseedor de la característica sea constante en todas la pruebas es necesario que las proporciones poblacionales no cambien tras cada extracción es decir se reemplace cada individuo extraído .Sin embargo si la población es muy grande, aunque no reemplacemos los individuos extraídos las variaciones en las proporciones de la población restante serán muy pequeñas y, aunque de hecho las probabilidades de, obtener un éxito varíen tras cada prueba, esta variación será muy pequeña y podremos considerar que son constantes .

En estadística, el test binomial es un test exacto de la significación estadística de las desviaciones de una teóricamente distribución esperada de las observaciones en dos categorías.

Uso común

Un uso común del test binomial es en el caso donde la hipótesis nula es aquella en la que las dos categorías son igualmente probables de que ocurran (como el lanzamiento de una moneda). Las tablas están ampliamente disponibles para dar la importancia observada en el número de observaciones en las categorías para este caso. Sin embargo, como muestra el siguiente ejemplo, el test binomial no se limita a este caso.

Donde hay más de dos categorías, y un test exacto es requerido, el test multinomial, basado en la distribución multinomial, debe usarse en vez del test binomial.

Muestras grandes

Para muestras grandes como las del siguiente ejemplo, La distribución binominal es bien aproximada por convenientes distribuciones continuas, y éstos se utilizan como la base para las pruebas alternativas que son mucho más rápidas para computar, Prueba χ^2 de Pearson y el G-test. Sin embargo, para pequeñas muestras estas aproximaciones se descomponen, y no hay alternativa para el test binomial.

La distribución de Poisson se aplica a varios fenómenos discretos de la naturaleza (esto es, aquellos fenómenos que ocurren 0, 1, 2, 3,... veces durante un periodo definido de tiempo o en un área determinada) cuando la probabilidad de ocurrencia del fenómeno es constante en el tiempo o el espacio. Ejemplos de estos eventos que pueden ser modelados por la distribución de Poisson incluyen:

El número de autos que pasan a través de un cierto punto en una ruta (suficientemente distantes de los semáforos) durante un periodo definido de tiempo.

El número de errores de ortografía que uno comete al escribir una única página.

El número de llamadas telefónicas en una central telefónica por minuto.

El número de servidores web accedidos por minuto.

El número de animales muertos encontrados por unidad de longitud de ruta.

El número de mutaciones de determinada cadena de ADN después de cierta cantidad de radiación.

El número de núcleos atómicos inestables que se han desintegrado en un determinado período.

El número de estrellas en un determinado volumen de espacio.

La distribución de receptores visuales en la retina del ojo humano.

La inventiva2 de un inventor a lo largo de su carrera.

3.5. Test basado en el estadístico chi-cuadrado.

Una prueba de chi-cuadrada es una prueba de hipótesis que compara la distribución observada de los datos con una distribución esperada de los datos.

Existen varios tipos de pruebas de chi-cuadrada:

Prueba de bondad de ajuste de chi-cuadrada

Utilice este análisis para probar qué tan bien una muestra de datos categóricos se ajusta a una distribución teórica.

Por ejemplo, usted puede comprobar si un dado es justo, lanzando el dado muchas veces y utilizando una prueba de bondad de ajuste de chi-cuadrada para determinar si los resultados siguen una distribución uniforme. En este caso, el estadístico de chi-cuadrada cuantifica qué tanto varía la distribución observada de los conteos con respecto a la distribución hipotética.

Pruebas de chi-cuadrada de asociación e independencia

Los cálculos para estas pruebas son iguales, pero la pregunta que se está tratando de contestar puede ser diferente.

Prueba de asociación: Utilice una prueba de asociación para determinar si una variable está asociada a otra variable. Por ejemplo, determine si las ventas de diferentes colores de automóviles dependen de la ciudad donde se venden.

Prueba de independencia: Utilice una prueba de independencia para determinar si el valor observado de una variable depende del valor observado de otra variable. Por ejemplo, determine si el hecho de que una persona vote por un candidato no depende del sexo del elector.

Esta prueba puede utilizarse incluso con datos medibles en una escala nominal. La hipótesis nula de la prueba Chi-cuadrado postula una distribución de probabilidad totalmente especificada como el modelo matemático de la población que ha generado la muestra.

Para realizar este contraste se disponen los datos en una tabla de frecuencias. Para cada valor o intervalo de valores se indica la frecuencia absoluta observada o empírica (Oi). A continuación, y suponiendo que la hipótesis nula es cierta, se calculan para cada valor o intervalo de valores la frecuencia absoluta que cabría esperar o frecuencia esperada (Ei=n ·pi, donde n es el tamaño de la muestra y pi la probabilidad del i-ésimo valor o intervalo de valores según la hipótesis nula). El estadístico de prueba se basa en las diferencias entre la Oi y Ei.

Este estadístico tiene una distribución Chi-cuadrado con k-1 grados de libertad si n es suficientemente grande, es decir, si todas las frecuencias esperadas son mayores que 5. En la práctica se tolera un máximo del 20% de frecuencias inferiores a 5.

Si existe concordancia perfecta entre las frecuencias observadas y las esperadas el estadístico tomará un valor igual a 0; por el contrario, si existe una grande discrepancia entre estas frecuencias el estadístico tomará un valor grande y, en consecuencia, se rechazará la hipótesis nula. Así pues, la región crítica estará situada en el extremo superior de la distribución Chi-cuadrado con k-1 grados de libertad.

El estadístico ji-cuadrado (o chi cuadrado), que tiene distribución de probabilidad del mismo nombre, sirve para someter a prueba hipótesis referidas a distribuciones de frecuencias. En términos generales, esta prueba contrasta frecuencias observadas con las frecuencias esperadas de acuerdo con la hipótesis nula. En este artículo se describe el uso del estadístico ji-cuadrado para probar la asociación entre dos variables utilizando una situación hipotética y datos simulados. Luego se describe su uso para evaluar cuán buena puede resultar una distribución teórica, cuando pretende representar la distribución real de los datos de una muestra determinada. A esto se le llama evaluar la bondad de un ajuste. Probar la bondad de un ajuste es ver en qué medida se ajustan los datos observados a una distribución teórica o esperada. Para esto, se utiliza una segunda situación hipotética y datos simulados.

Del mismo modo que los estadísticos "z", con su distribución normal y "t", con su distribución t de Student, nos han servido para someter a prueba hipótesis que involucran a promedios y porcentajes, el estadístico ji-cuadrado (o chi cuadrado), que tiene distribución de probabilidad del mismo nombre, nos servirá para someter a prueba hipótesis referidas a distribuciones de frecuencias.

En primer lugar, usaremos el estadístico ji-cuadrado para probar la asociación entre dos variables, y luego lo usaremos para evaluar en qué medida se ajusta la distribución de frecuencias obtenida con los datos de una muestra, a una distribución teórica o esperada.

En términos generales, esta prueba contrasta frecuencias observadas con las frecuencias esperadas de acuerdo con la hipótesis nula. Al igual que en el caso de las pruebas anteriormente presentadas, ilustraremos con ejemplos.

3.5.1. Test de bondad de ajuste.

La bondad de ajuste de un modelo estadístico describe lo bien que se ajusta un conjunto de observaciones. Las medidas de bondad en general resumen la discrepancia entre los valores observados y los valores esperados en el modelo de estudio. Tales medidas se pueden emplear en el contraste de hipótesis, e.g.

el test de normalidad de los residuos, comprobar si dos muestras se obtienen a partir de dos distribuciones idénticas (ver test de Kolmogorov-Smirnov), o si las frecuencias siguen una distribución específica (ver ji cuadrada).

Distribuciones

Para calcular si una distribución dada se ajusta a un conjunto de datos, se pueden utilizar las siguientes pruebas:

- Prueba de Kolmogórov-Smirnov
- Criterio de Cramér-von Mises
- Prueba de Anderson-Darling
- Test de Shapiro–Wilk
- Prueba de ji cuadrada
- Criterio de Información de Akaike

Regresiones

En el análisis de regresión, existen los siguientes indicadores:

- Coeficiente de determinación (El R2 mide la bondad de ajuste)
- Lack-of-fit sum of squares.

En los sistemas reales regularmente nos encontramos con variables cuyo comportamiento es aleatorio, y son susceptibles de ser modeladas por variables de entrada de un modelo estocástico, estás variables requieren de un tratamiento estadístico para su generación de manera artificial, el cual se realiza usualmente por medio de un modelo teórico de distribución de probabilidad, es así como las pruebas de bondad de ajuste es una buena herramienta para determinar el comportamiento de un conjunto de datos.

En muchas ocasiones cuando se está simulando un sistema, las variables son controlables del modelo son estocásticas, las variables de entradas las cuales tienen un comportamiento aleatorio son muestreadas con el objetivo detener un conjunto de datos sobre dicha variable aleatoria y encontrar el modelo de distribución de probabilidad que pueda representar las series de datos productos de la muestra.

En otras palabras, lo que se desea es probar la hipótesis que un modelo de probabilidad teórico (normal, exponencial, poissón etc.) en particular será un modelo satisfactorio de la población en estudio.

Este ajuste de los datos a un modelo de distribución de probabilidad se puede realizar por medio de las pruebas estadísticas más conocidas como pruebas de bondad de ajuste tales como la chi cuadrado y la de Kolmogor-smirnov.

Debe tenerse en cuanta que cuando a una serie de datos se la aplica cualquiera de las pruebas de bondad y se encuentra que ningún modelo teórico se puede ajustas a la serie de datos, se trabaja entonces con el modelo empírico (que no es modelo estándar conocido).

PRUEBA DE BONDAD DE AJUSTE CHI CUADRADO x2

El procedimiento de la prueba requiere una muestra aleatoria de tamaño n proveniente de la población cuya distribución de probabilidad es desconocida.

Estas n observaciones se pueden distribuir en k intervalos de clases y pueden ser representadas en histogramas.

La prueba se puede utilizar tanto para distribuciones discretas como para distribuciones continuas.

La prueba se puede sintetizar en los siguientes pasos.

1. Se colocan los n datos históricos (muéstrales) en una tabla de frecuencia.

2. Se propone una distribución de probabilidad una distribución de probabilidad de acuerdo con la tabla de frecuencia o con la curva que muestre un histograma o polígono de frecuencia.

3. Con la distribución propuesta, se calcula la frecuencia esperada para cada uno de los intervalos (FEi) de la siguiente manera:

   - Si la variable es continua se halla mediante la integración de la distribución propuesta y luego se multiplica por el número total de datos.
   - Si la variable es continua se utiliza de modelo matemático de la distribución propuesta y se evalúan todas las categorías y luego se multiplica por el número total de datos.

4. Se calcula el estadístico de prueba

5. Si el estimador C es menor o igual al valor correspondiente X2 con m-k-1 grados de libertad (K= números de parámetros estimados de la distribución propuesta estimada por los estadísticos muéstrales) y a un nivel de confiabilidad de 1-α, entonces no se puede rechazar la hipótesis de que los datos siguen la distribución que se propuso.

3.5.2. Test de heterogeneidad.

Cuando hablamos de heterogeneidad podemos distinguir dos aspectos: por un lado el relativo a las diferencias existentes entre los estudios en cuanto a características de los pacientes incluidos, la metodología utilizada, el tiempo de seguimiento, las dosis empleadas, la localización geográfica, etc. y por otro lado el concepto de heterogeneidad estadística ,que únicamente cuantifica la variabilidad entre los resultados de los estudios, y que puede ser debida a las diferencias reales de planteamiento y ejecución entre los estudios incluidos, o a otras causas.

La elaboración de guías de práctica clínica se fundamenta cada vez más en los resultados de revisiones sistemáticas, en las que los meta-análisis constituyen una herramienta crucial, y puesto que la validez de la conclusión global de un meta-análisis depende en gran medida de la homogeneidad de los estudios incluidos, resulta de extraordinaria importancia disponer de algún parámetro que cuantifique la heterogeneidad.

Uno de los aspectos de la heterogeneidad, el relativo a las diferencias clínicas o biológicas entre estudios y a las diferencias de procedimientos, es en primer lugar un problema metodológico, ya que habrá que decidir si las diferencias entre los estudios, que siempre existen, permiten o no combinarlos, independientemente de los resultados que en ellos se haya obtenido, y por lo tanto es previo a la ejecución del meta-análisis. Mientras que la heterogeneidad estadística trata de cuantificar la variabilidad del resultado medido en los diferentes estudios con respecto al resultado global promedio, y determinar si dicha variabilidad es superior a la que sería esperable por puro azar.

PRUEBA ESTADÍSTICA PARA VERIFICAR LA EXISTENCIA DE HETEROGENEIDAD

La prueba estadística más ampliamente utilizada para verificar la posible existencia de heterogeneidad superior a la esperable por puro azar se denomina Q de Cochran, y se basa en calcular la suma de las desviaciones cuadráticas entre el resultado individual de cada estudio y el resultado global, ponderadas por el mismo peso con el que cada resultado interviene en el cálculo global.

Sin embargo el empleo de esta prueba no está exento de problemas, ya que si el número de estudios es pequeño su capacidad para detectar heterogeneidad es muy baja (poca potencia de contraste), mientras que, por el contrario, cuando el meta-análisis combina gran número de estudios, el resultado puede ser estadísticamente significativo incluso cuando la magnitud de la heterogeneidad no sea de relevancia clínica. Esto no es nada nuevo: son los problemas inherentes a la metodología de las pruebas de contraste estadístico.

Puesto que la situación más habitual en los meta-análisis es la de un reducido número de estudios, se recomienda utilizar un nivel de rechazo en esta prueba al menos de 0.1, en lugar del tradicional 0.05.

Otras alternativas para cuantificar la heterogeneidad en un meta-análisis

Las limitaciones de la prueba Q anteriormente comentadas, así como el hecho de que su resultado no sirva para comparar diferentes meta-análisis, en los que intervienen diferente número de estudios, ha llevado a buscar otros índices que permitan cuantificar la heterogeneidad de manera que el parámetro calculado sirva para comparar diferentes meta-análisis y así, de alguna manera, ponderar la validez del efecto medio en ellos calculado. En esa línea se ha propuesto recientemente un nuevo índice denominado I2, que comienza a aparecer ya en las revisiones sistemáticas y del que se dispone de una muy buena descripción en un reciente artículo de BMJ.

Partiendo de la idea de que la pregunta de interés no es si existe o no heterogeneidad estadísticamente significativa, sino cómo afecta ésta a las conclusiones del meta-análisis, los autores trataron de buscar un índice que permita la comparación entre diferentes meta-análisis y por tanto que no dependa ni de la unidad de medida utilizada para cuantificar el efecto, ni del número de estudios incluidos, y que además sea fácilmente interpretable por personal no muy experto en estadística. Como veremos seguidamente, el índice propuesto cumple esos requisitos.

La heterogeneidad está relacionada con las diferencias que se observan entre los resultados de los estudios incluidos. Esta presentación muestra cómo explorar estas diferencias y cómo esto informará la comprensión de los efectos observados, así como cómo suelen interpretarse. Incluye una discusión de los modelos de metanálisis de efecto fijo y efectos aleatorios, que evalúa la heterogeneidad utilizando la prueba de ji cuadrado o el estadístico I2 y que explora la heterogeneidad mediante análisis de subgrupos.

3.5.3. Test de homogeneidad.

Se plantea el problema de la existencia de homogeneidad entre r poblaciones, para lo cual se realizan muestras independientes en cada una de ellas. Los datos muéstrales vienen clasificados en s clases y sus frecuencias absolutas se presentan en forma de una matriz r x s.

siendo nij el número de observaciones en la i-ésima población pertenecientes a la j-ésima clase.

Se quiere contrastar la hipótesis nula de que las probabilidades asociadas a las s clases son iguales en las r poblaciones. Donde es el tamaño muestral para la i-ésima población, es la frecuencia marginal de la j-ésima clase y es el tamaño muestral total. El estadístico L se distribuye como una con (r - 1)(s - 1) grados de libertad. El contraste se realiza con un nivel de significación del 5%.

Objetivo de la prueba: se utiliza cuando se tienen varias muestras independientes de n individuos que se clasifican respecto a una variable cualitativa y se desea conocer a partir de datos muéstrales, si provienen de la misma población (el objetivo es comparar diferentes muestras).

Es decir, en esta prueba se tienen varias muestras independientes correspondientes a las categorías de una de las variables y se clasifican las observaciones respecto a la otra variable. La prueba tiene la finalidad de conocer si la distribución de la variable estudiada difiere en las “r" poblaciones subyacentes de las cuales se obtuvieron las muestras.

Limitaciones de la prueba (las mismas que para la prueba de Independencia):

-Se necesita que más del 20 % de los valores esperados estén por encima de 5 y que ninguna celda tenga valor esperado menor a 1.

-Si la tabla es de 2x2, todas las celdas deben tener valores esperados por encima de 5.

-En el caso de la tabla de 2x2 si existe una sola celda con valor esperado menor que 5, esto representaría un 25 % de las celdas con esa condición, por lo que se utilizaría la Prueba de las Probabilidades exactas de Fisher en lugar de la Prueba χ 2, ya que en éste caso no es posible agrupar categorías.

El estadígrafo de prueba y la regla de decisión son similares a los de la Prueba Ji-cuadrado de independencia.

Errores más frecuentes en el uso de las Pruebas de Independencia y Homogeneidad:

• No inspeccionar los datos antes de realizar cualquier prueba de hipótesis.

• Clasificar una variable cuantitativa en su naturaleza con una escala de menor para poder realizar la prueba.

• Utilizar la prueba cuando una de las variables es cualitativa ordinal (en ese caso se emplea la Ji-cuadrado Tendencia Lineal).

• Usar el estadístico como una medida de asociación (estas pruebas son de significación de asociación y no dan una medida de asociación, solo permiten identificar si existe o no a asociación, pero no cuantifican la magnitud de esa asociación en caso de que exista).

• Usar la prueba cuando se dispone de valores promedios o porcentajes (la prueba solo puede realizarse con las frecuencias observadas, no con medidas de resumen).

• En las tablas FxC: no se debe utilizar la prueba cuando más del 20% de las celdas tienen frecuencias esperadas menores que 5 o al menos 1 de las celdas tiene frecuencia esperada inferior a 1.

• En las tablas 2x2: no se debe utilizar la prueba cuando una frecuencia esperada es menor que 5. En ese caso debe realizarse la Prueba de Fisher-Irwin o de probabilidad exacta de Fisher.

3.5.4. Tablas de Contingencia.

En estadística las tablas de contingencia se emplean para registrar y analizar la asociación entre dos o más variables, habitualmente de naturaleza cualitativa (nominales u ordinales).

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

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

Las tablas de contingencia más simples son tablas de dos factores que cuentan las respuestas según dos variables. Usted puede categorizar las observaciones según tres o más variables al "cruzarlas".

Una tabla de contingencia es una de las formas más comunes de resumir datos categóricos. En general, el interés se centra en estudiar si existe alguna asociación entre una variable denominada fila y otra variable denominada columna y se calcula la intensidad de dicha asociación.

De manera formal, se consideran X e Y dos variables categóricas con I y J categorías respectivamente. Una observación puede venir clasificada en una de las posibles $I \times J$ categorías que existen.

Cuando las casillas de la tabla contienen las frecuencias observadas, la tabla se denomina tabla de contingencia, término que fue introducido por Pearson en 1904.

Una tabla de contingencia (o tabla de clasificación cruzada), con I filas y J columnas se denomina una tabla I × J.

La tabla de contingencia es un medio particular de representar simultáneamente dos caracteres observados en una misma población, si son discretos o continuos reagrupados en clases. Los dos caracteres son el tamaño de la muestra.

UNIDAD IV

4.1. Conceptos de demografía.

La demografía es una ciencia que estudia las poblaciones humanas, su dimensión, estructura, evolución y características generales.

La demografía estudia estadísticamente la estructura y la dinámica de las poblaciones, así como los procesos concretos que determinan su formación, conservación y desaparición. Tales procesos son los de fecundidad, mortalidad y migración: emigración e inmigración. Para Livi Bacci, aclarado en buena medida el concepto de población se puede acometerse el de demografía:

A partir de la definición de población puede deducirse una definición de la demografía, la cual estudiaría aquellos procesos que determinan la formación, la conservación y la desaparición de las poblaciones. Tales procesos, en su forma más agregada, son los de fecundidad, mortalidad y movilidad. La variedad de combinaciones de estos fenómenos, interdependientes entre sí, supone la velocidad de las modificaciones de la población, tanto en sus dimensiones numéricas como en su estructura poblacional. Massimo Livi Bacci, Introducción a la demografía.

Es el estudio interdisciplinario de las poblaciones humanas. La demografía trata de las características sociales de la población y de su desarrollo a través del tiempo. Los datos demográficos se refieren, entre otros, al análisis de la población por edades, situación familiar, grupos étnicos, actividades económicas y estado civil; las modificaciones de la población, nacimientos, matrimonios y fallecimientos; esperanza de vida, estadísticas sobre migraciones, sus efectos sociales y económicos; grado de delincuencia; niveles de educación y otras estadísticas económicas y sociales.

La demografía es el análisis de las comunidades humanas a partir de la estadística. El concepto procede de un vocablo griego compuesto que puede traducirse como “descripción del pueblo”. Esta disciplina estudia el tamaño, la estratificación y el desarrollo de una colectividad, desde una perspectiva cuantitativa.

Para la demografía, la población es un conjunto de personas vinculadas por nexos reproductivos que puede identificarse por particularidades culturales, sociales, geográficas, políticas o de otro tipo. La población, por lo tanto, tiene continuidad en el tiempo, aunque no es eterna.

A partir de esta definición, podemos entender que la demografía es la ciencia social que se encarga de analizar los procesos que fijan la creación, el mantenimiento y la eventual desaparición de las poblaciones. Nociones como fecundidad, movilidad y mortalidad son claves para la demografía, ya que determinan la estructura de cada población.

El estadista y sociólogo árabe Ibn Jaldún (1332–1406) está considerado como el pionero en el ámbito de la demografía. Él fue quien comenzó a recopilar información estadística para estudiar a las poblaciones y generar nuevos datos a partir de estas estadísticas. Los británicos John Graunt (1620–1674) y Thomas Malthus (1766–1834) también hicieron grandes aportes al desarrollo de la demografía.

Las tasas de natalidad y mortalidad y los movimientos de la población forman parte de los estudios básicos de la demografía, cuyos trabajos resultan muy importantes para el desarrollo de las políticas de Estado. A partir de los datos demográficos, por ejemplo, se puede conocer cuáles son los principales problemas de la salud pública o qué regiones de un país se encuentran relegadas a nivel económico.

Dentro de esta ciencia social existen diversas ramas, las mismas responden a las siguientes nombres: demografía general (investiga las teorías que existen en torno a la demografía y las metodologías de investigación utilizadas), demografía geográfica (chequea la movilidad de las poblaciones: migraciones, nuevos asentamientos, etc), demografía histórica (estudia cuestiones relacionadas con la fertilidad, índice de mortalidad y las migraciones que se dan dentro de un grupo), fertilidad (que se encarga de analizar los índices de natalidad y matrimonio y la fecundidad de la población) y mortalidad (estudia el índice de mortalidad de un grupo en general y las causas y edad de las muertes en particular, intentando relacionar las diversas variables).

Para realizar los estudios demográficos se llevan a cabo diversos censos que tienen por objetivo extraer la información relevante en torno al estado de la población que se está estudiando: número de viviendas ocupadas, cantidad de personas que viven en cada una, enfermedades, muertes acaecidas en el último año en cada familia, migraciones, etc.

Los resultados de estos análisis no sólo permitirán conocer el tamaño actual de la población, sino también los riesgos a las que se encuentran expuestos y puede ayudar a buscar soluciones o prevenir consecuencias letales como pestes, hambrunas o accidentes.

4.1.1. Conceptos básicos.

Características generales

La ciencia de la demografía no se limita a la medición, sino que incluye necesariamente la interpretación y análisis de los datos, las proyecciones y previsiones en base a supuestos que incluyen variables no demográficas. Sin embargo, la demografía estadística es el punto de partida del análisis de la población en el que se trata de medir con precisión las magnitudes demográficas.

Fecundidad

El concepto de fecundidad se refiere al número medio de hijos que tienen las mujeres. Para medirlo con precisión es necesario delimitar con precisión la variable que se pretende medir ya que la cifra que la exprese será muy distinta según se considere a todas las mujeres que viven en un momento determinado en un país, o sólo a las mujeres fértiles, eliminando las que mueren antes de alcanzar la edad fértil. Se puede estimar también tasas de fecundidad por edades o tasa de fecundidad de cohortes.

Crecimiento demográfico

El crecimiento demográfico mide el aumento, en un período específico, del número de personas que viven en un país o una región. La tasa de crecimiento demográfico depende, además de la tasa de natalidad y de la tasa de mortalidad, de los movimientos migratorios. La tasa de natalidad depende a su vez de la tasa de fecundidad. La tasa de fecundidad está influida por muchos factores, pero el principal es el nivel cultural de la sociedad y especialmente de las mujeres: a mayor cultura, menor número de hijos se tienen. La tasa de mortalidad depende del grado de desarrollo económico y sanitario.

Longevidad

La longevidad es la duración de la vida de una persona. Se mide mediante el concepto de esperanza de vida. La esperanza de vida de un tipo de persona es la media de la duración de la vida de ese tipo de personas. También podemos calcular la esperanza de vida a los 75 años en 1963: cuánto tiempo sobrevivieron de media las personas que ese año tenían una edad de 75.

Evolución de la población mundial

Los ecologistas han estimado que la Tierra pudo proporcionar a las bandas de cazadores-recolectores alimento suficiente para un máximo de treinta millones de individuos. En los cuatro millones de años que requirió la evolución desde el "homo erectus" al hombre actual, no se pudo superar esa cifra. Posiblemente la población total del Paleolítico oscilaría entre los seis y los diez millones de seres humanos.

La revolución neolítica, hace diez mil años, mediante la aplicación de técnicas agrícolas y ganaderas permitió la primera gran expansión de la especie humana.

Se calcula que a partir de entonces la población empezó a crecer a un ritmo que la duplicaba cada mil setecientos años. Al comienzo de nuestra era se calcula que vivían unos ciento cincuenta millones de personas: una tercera parte en el Imperio Romano, otra tercera parte en el Imperio Chino y el resto diseminado.

Importancia

- Elaboración de tasas y otros indicadores de salud
- Estudios en epidemiología: En los estudios epidemiológicos se necesitan datos de la población y de su distribución según características de persona, lugar y tiempo.
- Planificación de la salud pública
- Planificación de producción alimentaria (alimentación humana)
- Planes generales de desarrollo nacionales o regionales
- Proyecciones de las poblaciones para cualquier propósito

Tipos de demografía

Los dos tipos o partes de la demografía están interrelacionados entre sí, y la separación es un tanto artificial, puesto que el objetivo de estudio es el mismo: las poblaciones humanas.

Demografía estática: Es la parte de la demografía que estudia las poblaciones humanas en un momento de tiempo determinado desde un punto de vista de dimensión, territorio, estructura y características estructurales.

- La dimensión es el número de personas que residen normalmente en un territorio geográficamente bien delimitado.

- El territorio es el lugar de residencia de las personas que puede globalizarse o desagregarse como, por ejemplo, una nación, una región, una provincia, una ciudad, un municipio, etc.
- La estructura de una población es la clasificación de sus habitantes según variables de persona. Según las Naciones Unidas, estas variables son: edad, sexo, estado civil, lugar de nacimiento, nacionalidad, lengua hablada, nivel de instrucción, nivel económico y fecundidad.

Demografía dinámica: Es la parte de la demografía que estudia las poblaciones humanas desde el punto de vista de la evolución en el transcurso del tiempo y los mecanismos por los que se modifica la dimensión, estructura y distribución geográfica de las poblaciones. Ejemplos de tales mecanismos son la natalidad, la mortalidad, la familia, la fecundidad, el sexo, la edad, la educación, el divorcio, el envejecimiento, la migración, el trabajo, la emigración y la inmigración.

Su expresión son las tablas demográficas, que son los datos estadísticos numéricos y gráficos. Los administradores utilizan el censo total (real (cada diez años) o muestreos (estimados) mensuales o anuales). Tienen que estar diseñados de forma que no sólo se puedan establecer estadísticas descriptivas, sino que también puedan realizarse análisis demográficos cruzados. Unos ejemplos son: los sueldos anuales y las edades o la distribución por sueldos, por clases sociales, en la nación y en una comunidad, por viviendas y familias, etc. Tópicos generales o variables básicas son la riqueza, el poder y la movilidad social.

Los datos estadísticos sobre las poblaciones también son sometidos a análisis predictores o de futuro: interpolaciones, extrapolaciones, series de tiempo, curvas logísticas, patrones de crecimiento según tipo de sociedad, patrones de disminución por desastres naturales o epidemias o guerras, etc.

Desde el siglo XIX se descubrió que la gráfica del crecimiento de las poblaciones sigue la forma de una S alargada, de crecimiento rápido o modelo exponencial, llega a un punto de inflexión y continúa con un crecimiento suave, y es un reflejo del paso de una sociedad agrícola a una sociedad industrial: la reducción en el número de nacimientos y el aumento en la población que se halla en la tercera edad.

Teorías demográficas:

Transición demográfica

Según el modelo de la transición demográfica los cambios en una población tienen tres componentes: nacimientos, muertes y migración. La transición demográfica como proceso, disminuye el efecto de cuatro grandes riesgos sociodemográficos: la alta mortalidad, la elevada fecundidad, el acelerado crecimiento demográfico y la estructura etaria joven.

Con mayor o menor intensidad y velocidad, todos los países de América Latina han iniciado el proceso de la transición demográfica, porque todos han empezado a reducir sus tasas de fecundidad y mortalidad.

El Celade ha elaborado topologías para identificar los riesgos sociodemográficos por los que atraviesan los países, según el grado de avance de la transición demográfica. La transición demográfica se caracteriza por diferentes fases.

La transición incipiente: con alta natalidad y mortalidad y con un crecimiento natural moderado, del orden de 2.5%, tienen una estructura por edades muy joven y una alta relación de dependencia.

La transición moderada: alta natalidad, pero cuya mortalidad es moderada. Por este motivo su crecimiento natural es todavía elevado, cercano al 3%. En esta etapa se ubica, por ejemplo, Guatemala, donde el descenso de la mortalidad, sobre todo durante el primer año de vida, se ha traducido en un rejuvenecimiento de la estructura por edades, lo que también lleva a una elevada relación de dependencia.

La plena transición: con natalidad moderada y mortalidad moderada o baja, lo que determina un crecimiento natural moderado cercano al 2%. Aquí el descenso de la fecundidad es reciente y la estructura por edades se mantiene todavía relativamente joven, aun cuando ya ha disminuido la relación de dependencia.

La transición avanzada: con natalidad y mortalidad moderada o baja, lo que se traduce en un crecimiento natural bajo, del orden del 1%.

Segunda transición demográfica

El concepto de la segunda transición demográfica fue creado por Lesthaghe y D.J. van de Kaa en 1986.5 Es un concepto nuevo que procura dar cuenta de fenómenos emergentes en países desarrollados, pero que también parece que se confirma en países de América Latina.

La segunda transición demográfica, en un contexto estable de baja fecundidad y mortalidad, describe los cambios en la composición de la familia y de las uniones en los patrones de reconstitución de las familias en países occidentales.

Además de niveles de fecundidad inferiores al nivel de reemplazo y sostenidos en el tiempo, la segunda transición demográfica se caracteriza por: (i) incremento de la soltería, (ii) retraso del matrimonio, (iii) postergación del primer hijo, (iv) expansión de las uniones consensuales, (v) expansión de los nacimientos fuera de matrimonio, (vi) alza de las rupturas matrimoniales, (vii) diversificación de las modalidades de estructuración familiar.

Revolución reproductiva

Artículo principal: Revolución reproductiva

La teoría de la revolución reproductiva es crítica a las limitaciones de la teoría general de la transición demográfica derivadas de su metodología de investigación -apoyada en estudios transversales y expresada en las pirámides de población- ya que proyectaría una visión incompleta no holística de la dinámica población (nuevas fases se deben incorporar a la transición demográfica para dar cuenta de nuevos fenómenos) dejando sin explicación algunos de los mecanismos de reproducción de las poblaciones en la sucesión intergeneracional que se están manifestando en las sociedades modernas. Como propuesta de cambio de paradigma la revolución reproductiva -apoyada en estudios longitudinales- pretende dar cuenta de los cambios demográficos de manera sistémica y no alarmista ni catastrofista, integrando en buena medida las consecuencias sociológicas que caracterizan la segunda transición demográfica.

La teoría de la revolución reproductiva está descrita por John MacInnes8 y Julio Pérez Díaz en sus publicaciones The reproductive revolution de 2005 y de 2009 La tercera revolución de la modernidad. La revolución reproductiva y The reproductive revolution. Los autores señalan la radical relevancia que tiene en su teoría y en los nuevos fenómenos demográficos el concepto de eficiencia reproductiva, así como la longevidad y el reemplazo generacional en las sociedades modernas. Siguiendo el hilo conductor de las ideas de Kingsley Davis (1908-1997) expuestas en 1937 sobre el futuro de la familia y de la fecundidad, establece consecuencias muy distintas sobre las implicaciones y consecuencias que la revolución reproductiva tiene en el descenso del trabajo reproductivo: el declive del patriarcado, la desregulación social de la sexualidad, el paso del género a la generación como eje de distribución de roles productivos-reproductivos, el reforzamiento de lazos familiares y otras consecuencias positivas de la madurez de masas el mal llamado envejecimiento de la población.

Variación de edades y sexo

En la mayoría de países del planeta la población de las mujeres es mayor a la de los hombres, aunque en unos pocos países como Andorra, Albania, China, Costa Rica, Filipinas, India, la mayor parte de los países de Oriente Medio, Panamá, Paraguay y República Dominicana entre otros, se estima con una población masculina mayoritaria.

Sin embargo, el caso de los países donde la mayoría son personas de sexo femenino, es porque se incluye la ancianidad en las estimaciones. Las personas de sexo femenino de la población de la tercera edad son mayoría en todo el mundo. Esto está de acuerdo con la ciencia que establece que la mujer disfruta de más longevidad que los hombres. No obstante, si dejamos esta etapa a un lado encontraremos que la población de niños, adolescentes y adulta muchas veces los varones son mayoría, por ejemplo, en países como Alemania, Francia, Japón, Corea del Sur, Cuba, Bélgica, España, Italia, Reino Unido, etc. En estos últimos años los niños y adolescentes en la población de sexo masculino superan a las personas de sexo femenino, pero en la población joven y adulta las personas de sexo femenino siguen siendo mayoría, por ejemplo, en países como Argentina, Bolivia, Brasil, Canadá, El Salvador, Estados Unidos, Filipinas, Guatemala, Guinea Ecuatorial, Guinea Bissau, Perú, Portugal, Puerto Rico, Rusia y algunos países de los Balcanes entre otros.

Además, se estima que en algunos países como Bélgica, Canadá, Cuba, Francia, Israel, Japón, Puerto Rico algunos países árabes y entre otros, la población masculina podría igualar y superar a la población femenina, ya que en estos países existe un franco crecimiento.

Importancia de la demografía en la salud pública

- Elaboración de tasas y otros indicadores de salud
- Estudios en epidemiología: En los estudios epidemiológicos se necesitan datos de la población y de su distribución según características de persona, lugar y tiempo.
- Planificación de la salud pública
- Planificación de producción alimentaria (alimentación humana)
- Planes generales de desarrollo nacionales o regionales
- Proyecciones de las poblaciones para cualquier propósito

Recursos demográficos en países de habla hispana

- Encuestas
- Padrones
- Registro Civil e Instituto Nacional de Estadística en España
- En Argentina, el Instituto Nacional de Estadística y Censos.
- En Bolivia, el Instituto Nacional de Estadística.
- En Chile, el Instituto Nacional de Estadísticas

- En Colombia, el Departamento Administrativo Nacional de Estadística.
- En Ecuador, el Instituto Nacional de Estadística y Censos de Ecuador.
- En España, el Instituto Nacional de Estadística.
- En México, el Instituto Nacional de Estadística y Geografía.
- En Paraguay, la Dirección General de Estadísticas, Encuestas y Censos.
- En el Perú, el Instituto Nacional de Estadística e Informática.
- En Puerto Rico, el Registro Demográfico de Puerto Rico, el Instituto de Estadísticas de Puerto Rico y la Oficina del Censo - Junta de Planificación de Puerto Rico.
- En Uruguay, el Instituto Nacional de Estadística de Uruguay.
- En Venezuela, el Instituto Nacional de Estadística de Venezuela.
- En Nicaragua, el Instituto Nacional de Estadística de Nicaragua.
- En Panama, el Instituto Nacional de Estadística y Censo

Métodos de estudio de la demografía

Existen dos tipos de métodos de estudio dentro de esta ciencia social:

Método compuesto: Se trata de combinar diferentes técnicas de estudios que permiten arribar a una conclusión aproximada de las condiciones en las que se encuentra la población estudiada. Una de estas técnicas puede ser por ejemplo, la matrícula escolar. Los investigadores pueden tener una noción aproximada de los cambios que ha sufrido la población de un año a otro en el sector del grupo que se encuentra en edad escolar, de este modo podrían conocer la cantidad de habitantes menores que hay y comprobar las migraciones que ha sufrido la población en el último año de jóvenes en edad escolar.

Métodos estadísticos: Son los más exactos porque se realizan a partir de la obtención de datos específicos, recogidos de los censos. A través de teorías estadísticas, se pueden relacionar los cambios que se indican en los resultados de los censos y obtener información sobre las condiciones en las que se encuentra la población al momento de realizar dicho análisis.

Demografía es, a grandes rasgos, la disciplina social que estudia estadísticamente a la población humana.

Diferentes definiciones de demografía

Demografía. Estudio estadístico de una colectividad humana, referido a un determinado momento o a su evolución.

La demografía es la ciencia que tiene por objeto el estudio de las poblaciones humanas tratando, desde un punto de vista principalmente cuantitativo, su dimensión, su estructura, su evolución y sus características generales.

La demografía estudia aquellos procesos que determinan la formación, la conservación y la desaparición de las poblaciones. Tales procesos, en su forma más agregada, son los de fecundidad, mortalidad y movilidad.

La demografía es una ciencia cuyo objeto es el hombre considerado en la totalidad de los aspectos de su realidad: como miembro de una colectividad a la que ingresa por el solo hecho de nacer y de la que se retira cuando muere. Esta realidad tiene diferentes ángulos. El hombre objeto de la demografía es un ser vivo y complejo, esto es: social, político, histórico, económico y moral. En este sentido puede decirse que la demografía es una ciencia antropológica, pero no un capítulo de la antropología, pues considera al hombre en colectividad, no en forma individual.

El análisis demográfico se refiere al conocimiento del comportamiento de los componentes de la población: la natalidad, la mortalidad y la migración, así como a sus cambios y consecuencias; a los factores que determinan los cambios y al periodo de tiempo requerido para que ocurran esos cambios. Los estudios de población se ocupan de las relaciones que existen entre los cambios de población y otros tipos de variables sociales, económicas, políticas, biológicas, genéticas y geográficas.

¿Qué es población?

Comúnmente la palabra población se emplea para designar el conjunto de los habitantes de un territorio determinado, como también para designar una parte de dicha población (p.ej., población en edad escolar, población en edad de casarse) aunque en este caso resulta más apropiado hablar de subpoblación. Con frecuencia se habla de población no para indicar el conjunto mismo sino el número de habitantes que la componen.

Por población se entiende un conjunto de individuos, constituido de forma estable, ligado por vínculos de reproducción e identificado por características territoriales, políticas, jurídicas, étnicas o religiosas. El significado de población es bastante elástico; este concepto abarca tanto pequeños grupos de algunos centenares de personas aislados por motivos geográficos, religiosos, etc., que a pesar de sus exiguas dimensiones consiguen asegurar su propia reproducción y supervivencia, como grandes naciones con varios centenares de millones de habitantes.

¿Cuáles son los elementos básicos de la demografía?

Lara y Mateos cita los siguientes:

Volumen. Número de personas de una población.

Distribución. Ordenamiento de la población en el espacio en un momento dado, que puede ser geográficamente, o entre varios tipos de áreas residenciales.

Estructura. Distribución de la población entre grupos de edad y sexo.

Cambio. Crecimiento o declinación de la población total o en algunas de sus unidades estructurales, los componentes esenciales de cambio son nacimientos, defunciones y migraciones.

Fuentes de datos en demografía

Básicamente hay dos tipos de fuentes:

Las que se basan en empadronamientos. Censos de población y encuestas demográficas.

Las que se basan en registros. Registros administrativos y estadísticas vitales.

Tipos y disciplinas

Lara y Mateos relaciona los siguientes tipos de demografía:

Demografía estática. Estudio del número absoluto de individuos que constituye una población determinada, repartidos en categorías con arreglo a su estado, edad, sexo, profesión, condiciones intelectuales, etc., y las relaciones que existen entre las diversas categorías.

Demografía dinámica. Estudia los movimientos internos que provienen de la natalidad y mortalidad, y los externos que tienen su origen en la inmigración y en la emigración.

Demografía general. Deduce de los datos anteriores las leyes o principios a que obedece la población y sus variaciones.

Asimismo, identifica las siguientes disciplinas dentro del campo de acción de la demografía:

Demografía descriptiva. Trata del volumen, distribución geográfica, estructura y desarrollo de las poblaciones humanas, apoyándose principalmente en las estadísticas demográficas, que es la aplicación de la estadística general al estudio de las poblaciones humanas.

Demografía teórica. Llamada también demografía pura, considera las poblaciones desde el punto de vista general y abstracto, estudiando las relaciones formales entre los distintos fenómenos demográficos.

Demografía cuantitativa. Se le llama así por la importancia que se le atribuye al aspecto numérico de los fenómenos y para distinguirla de las ramas que se expresan a continuación.

Demografía económica. Es la rama que trata de las poblaciones en relación con los fenómenos económicos.

Demografía social. Es la parte que se refiere a los fenómenos sociales.

Demografía y marketing

El entorno demográfico es una de las variables que incide en el macroentorno de marketing, siendo considerado de gran importancia dado que estudia a la población que conforma los mercados. Las tendencias y sucesos demográficos son analizados por las firmas para prever posibles comportamientos de mercados y consumos a futuro y para entender las causas de las acciones de marketing del pasado.

Por ejemplo, una variación en la estructura de edad en la población que constituye un mercado puede cambiar por completo el ciclo de vida del producto, obligando a las empresas a mantener activos sus departamentos de investigación y desarrollo de tal manera que los consumidores siempre estén satisfechos.

También interviene la demografía en el proceso de segmentación de mercados, ya que a través de ella se establecen las características de un determinado mercado en cuanto a edad, sexo, raza, localización y otras de las cuales se encargan el análisis y los estudios demográficos.

Demografía y macroeconomía

La demografía económica se constituye en pieza fundamental del análisis macroeconómico ya que a través de ella se evidencian los comportamientos y características de las poblaciones tanto transversalmente (en fechas específicas) como longitudinalmente (a través del tiempo).

4.1.2. Modelos de crecimiento de poblaciones.

Es necesario diferenciar entre tasa de crecimiento poblacional y tasa de crecimiento per-cápita.

Crecimiento poblacional independiente de la densidad

Presupuesto de estos modelos: La razón de mortalidad y natalidad per cápita no dependen del tamaño poblacional. Por lo tanto, la tasa de crecimiento per-cápita es constante,

Consecuencia de los modelos: La tasa de crecimiento poblacional es proporcional al tamaño poblacional.

Puntos más importantes:

En el crecimiento exponencial, la tasa de crecimiento per cápita (por individuo) de una población es la misma sin importar el tamaño de la población, lo que hace que crezca cada vez más rápido conforme se hace más grande.

En la naturaleza, las poblaciones pueden crecer de manera exponencial por un tiempo, pero finalmente se ven limitadas por la disponibilidad de recursos.

En el crecimiento logístico, la tasa de crecimiento per cápita se reduce cada vez más conforme el tamaño poblacional se acerca a un máximo impuesto por los recursos limitados del entorno, conocido como capacidad de carga (K).

El crecimiento exponencial produce una curva en forma de J, mientras que el crecimiento logístico produce una curva en forma de S.

Modelado de tasas de crecimiento

Para entender los diferentes modelos que se usan para representar las dinámicas poblacionales, empecemos por la ecuación general de la tasa de crecimiento poblacional (el cambio en el número de individuos en una población en el tiempo).

En esta ecuación, dN/dT, T es la tasa de crecimiento de la población en un momento determinado, N es el tamaño de la población, T es el tiempo, y r es la tasa de aumento per cápita, esto es, qué tan rápido crece la población por cada individuo que existe dentro de la misma.

Si suponemos que no hay un movimiento de individuos hacia adentro o hacia afuera de la población, entonces r es solo una función de las tasas de nacimiento y mortalidad.

La ecuación anterior es muy general y podemos hacer formas más específicas de ella para describir dos tipos diferentes de modelos de crecimiento: exponencial y logístico.

Cuando la tasa de aumento per cápita (r) toma el mismo valor positivo sin importar el tamaño de la población, entonces tenemos un crecimiento exponencial.

Cuando la tasa de aumento per cápita (r) disminuye a medida que la población alcanza su límite máximo, entonces tenemos un crecimiento logístico.

Crecimiento exponencial

Las bacterias cultivadas en el laboratorio son un excelente ejemplo de crecimiento exponencial. En el crecimiento exponencial, la tasa de crecimiento de la población aumenta con el tiempo, en proporción con el tamaño de la población.

El concepto fundamental del crecimiento exponencial es que la tasa de crecimiento poblacional el número de organismos que se añade en cada generación aumenta al mismo tiempo que la población se hace más grande. Los resultados pueden ser dramáticos: después de 1 día (24 ciclos de división) nuestra población de bacterias habría aumentado de 1000 ¡a más de 16 mil millones! Cuando se grafica el tamaño de la población N en el tiempo, se obtiene una gráfica en forma de J.

Crecimiento logístico

El crecimiento exponencial no es una situación muy sostenible, ya que depende de cantidades infinitas de recursos (las cuales no suelen existir en el mundo real).

El crecimiento exponencial puede ocurrir durante un tiempo, si hay pocos individuos y muchos recursos, pero cuando el número de individuos es lo suficientemente grande, los recursos empiezan a agotarse, lo que desacelera la tasa de crecimiento. Finalmente, el tamaño de la población se nivelará, o se estabilizará, lo que produce una gráfica con forma de S. El tamaño de la población en el que el crecimiento poblacional se nivela representa el tamaño poblacional máximo que puede soportar un medio ambiente en particular y se conoce como capacidad de carga o K.

Podemos modelar matemáticamente el crecimiento logístico al modificar nuestra ecuación del crecimiento exponencial usando una r (tasa de crecimiento per cápita) dependiente del tamaño poblacional (N) y de su cercanía a la capacidad de carga (K).

Cuando la población es pequeña, N es muy pequeña en comparación con K. En este punto el término (K - N)/K es aproximadamente (K/K), o 1, lo que resulta nuevamente en la ecuación del crecimiento exponencial.

MODELOS DE CRECIMIENTO

Modelo 1: Crecimiento Exponencial.

El primer modelo representa el crecimiento de la población en una fuente de presión constante. La fuente de presión constante puede abastecer tanta energía como se necesita. Por ejemplo, piense en una población de conejos en crecimiento, con abastecimiento de alimento que no considera la rapidez con que ellos comen. Siga el flujo del diagrama para ver cómo la población de conejos aumenta, esta retroalimenta para traer más energía (a través de más alimentación) para procrear más conejos. Si el sistema comienza con un conejo macho y una hembra, y ellos producen cuatro conejitos que a su vez producen ocho; y así, en la misma tasa de aumento, la próxima generación producirá 16, la próxima 32, la próxima 64 y así sucesivamente. Como el número de conejos aumenta, ellos usan más de la fuente de energía y el número aumenta rápidamente.

En la práctica, la fuente de energía a presión constante no puede ser mantenida indefinidamente, entonces el crecimiento exponencial infinito es imposible. De cualquier manera, durante las primeras etapas del crecimiento de la población, cuando la demanda de alimento es pequeña (comparada con la cantidad disponible) la energía puede estar disponible a presión constante y el crecimiento puede ser exponencial. Pero eventualmente, el alimento podría volverse limitante y la situación necesitaría ser representada por un modelo diferente.

Modelo 2: Crecimiento Logístico.

Las poblaciones creciendo inicialmente rápido en una fuente de presión constante, se vuelven tan numerosas que pierden su capacidad de crecer debido a interacciones entre los miembros de la población, resultando entonces un estado de equilibrio. Este tipo de crecimiento se llama crecimiento logístico.

Crecimiento logístico es el balance entre producción en proporción a la población, y a las pérdidas en proporción a la oportunidad de interacciones individuales.

El proceso de crecimiento puede ser ejemplo el crecimiento de levadura en el fermento del pan. Primeramente, el crecimiento de la población es casi exponencial. La disponibilidad de alimento es constante y como la población crece esto implica comer más y más. Sin embargo, las células de levaduras se vuelven tan numerosas que sus productos comienzan a interferir con el propio crecimiento. Resultando un estado de equilibrio entre producción y pérdida de células.

Modelo 3: Crecimiento en una fuente de flujo constante.

Los ecosistemas utilizan muchas fuentes cuyo flujo es controlado por sistemas externos. Ejemplos de fuentes de flujo constante son el sol, la lluvia, el viento y las corrientes de ríos. Las poblaciones en los sistemas no pueden aumentar los flujos externos. Su crecimiento se limita a aquello que pueda ser mantenido por el flujo interno de energía. Un ejemplo es la utilización de la luz solar por los árboles, no hay nada que los árboles puedan hacer para aumentar o disminuir la incidencia de luz solar. Este tipo de fuente es también llamado fuente renovable.

Un importante ejemplo en la naturaleza es la sucesión, como el crecimiento de una vegetación. Cuando la vegetación es joven, la energía de la luz no es limitante. El crecimiento de árboles pequeños es rápido y la mayoría del excedente de luz que pasa no es utilizada. Con el crecimiento de la vegetación, no obstante, los árboles utilizan más y más energía, y menos energía escapa de no ser utilizada. El crecimiento decrece y se detiene. La vegetación se vuelve un equilibrio entre crecimiento y descomposición.

Otro ejemplo de crecimiento, en una fuente de flujo constante, es la construcción de ciudades a lo largo de un río. Las ciudades usan agua para beber, producción agrícola, pesca y uso de aguas servidas tratadas. Nuevas ciudades pueden construirse hasta que toda el agua sea utilizada tan rápido cuanto fluya por el río.

4.1.3. Fuentes Históricas. Fuentes Naturales.

Las fuentes históricas constituyen la materia prima de la Historia. Comprenden todos los documentos, testimonios u objetos que nos trasmiten una información significativa referente a los hechos que han tenido lugar, especialmente en el pasado. Dentro de ellas, y considerando el valor que también tienen las demás, las Fuentes escritas son el apoyo básico para construir la Historia.

El historiador trabaja las fuentes históricas (las interroga y contrasta) para obtener de ellas la mayor información posible. Así mismo debe atender a su variedad, realizando una adecuada selección de las mismas. En sentido general, las Fuentes históricas son de dos tipos: primarias y secundarias.

- Fuentes primarias. Son las que se han elaborado prácticamente al mismo tiempo que los acontecimientos que queremos conocer. Llegan a nosotros sin ser transformadas por ninguna persona; es decir, tal y como fueron hechas en su momento, sin ser sometidas a ninguna modificación posterior.

- Fuentes secundarias. Se denominan también historiográficas. Son las que se elaboran a partir de las Fuentes primarias: libros, artículos...

La utilización de las fuentes. Metodología.

Para la confección del conocimiento histórico, las fuentes que utiliza el historiador deben ser analizadas, valoradas e interpretadas, siguiendo una metodología coherente. Además, el historiador debe tener encuentra las Fuentes en su momento histórico y en relación con las circunstancias en que surge o se elaboran. Deben ser sometidas a una crítica objetiva para conocer los elementos que las componen y comprobar su veracidad. Para ello el historiador utiliza un método, que consiste esencialmente en formular preguntas sobre su contenido, a partir de hipótesis de trabajo que queremos contrastar; el objetivo de este proceso es la construcción de la Historia.

El estudio de las fuentes históricas.

La metodología pormenorizada que proponemos a continuación responde básicamente a estos apartados: clasificación, explicación causal, intencionalidad, circunstancias históricas, análisis, comentario y aplicación de su contenido, utilidad y valoración general.

Las fuentes primarias: ¿Cómo afrontar su estudio?

Entendemos que estudiar las fuentes históricas en este nivel debe ser un procedimiento básico, atractivo y claro que nos permita identificarnos con la importancia y el contenido de un documento histórico. Para ello podemos seguir orientativamente los pasos siguientes:

a) Precisar que el documento nos informa sobre hechos y sucesos. Así realizaremos las preguntas siguientes: ¿qué ocurrió? Identificamos los hechos históricos. ¿Cómo sucedió? Realizamos su descripción. ¿Dónde?, ¿cómo?, ¿por qué?. ¿qué consecuencias están presentes en él?

b) El documento, ¿nos informa de grupos sociales o de personas?: ¿A quiénes se refiere?, ¿qué se dice de ellos?, ¿qué opinan, en su caso, las personas o grupos?

c) ¿Se informa en el documento acerca de diversas actividades?: precisar si son políticas, sociales, económicas, culturales, religiosas, etc. ¿Qué referencias hay de ellas? ¿Se precisa por qué se realizan?

d) ¿Informa el documento sobre instituciones?: ¿Cuáles?, ¿de qué tipo?, ¿qué función tienen?, ¿cómo están estructuradas?, ¿con quiénes se relacionan?

e) ¿Aporta datos concretos?: ¿Cómo se clasifican?, ¿de qué tipo son?, ¿tienen relación con personas, hechos, actividades, otras instituciones, etc.?

f) ¿Contiene opiniones significativas?: ¿de qué tipo?, ¿a qué o a quiénes hace referencia?, ¿qué actitudes reflejan?.

Las fuentes secundarias: ¿cómo afrontar su estudio?

Para su tratamiento y estudio, proponemos un procedimiento similar al que utilizamos en el comentario de textos históricos:

a) Lectura comprensiva e información previa: lectura atenta individual o en grupo. Localizar el vocabulario desconocido y buscar su significado. Subrayado de las ideas principales. Identificación de los temas tratados.

b) Análisis y clasificación: identificación de la naturaleza del texto (tipo de fuente) y su justificación. Señalización de los temas y problemas tratados, identificando sus ideas básicas. Delimitación del vocabulario histórico específico.

¿Qué sabemos del autor, destinatario y sus circunstancias? Situación del texto en su realidad histórica, y en las variables espacio-tiempo.

c) Comentario e interpretación: comentario de los temas analizados e ideas más significativas. Análisis de los hechos históricos, instituciones, personajes, etc., que aparecen y con los que se relaciona la fuente. Características de la época a la que se alude. Explicación de los antecedentes, causas y consecuencias que fundamentan su explicación. Valoración de su importancia para obtener información histórica.

d) Conclusión: síntesis del comentario realizado y opinión personal objetiva y fundamentada históricamente, basada en las aportaciones anteriores.

¿Y qué son las fuentes históricas?

Pues no son ni más ni menos que aquellos documentos, testimonios u objetos que nos trasmiten una información significativa referente a los hechos que han tenido lugar a lo largo del largo tiempo que nos observa y que el ser humano ha sido capaz de transcribir, bien de forma escrita, bien de forma oral.

Las fuentes históricas son el material sobre el que el historiador estudia y llega a conclusiones, siempre contrastándolas y poniéndolas en conocimiento de otros historiadores para generar el necesario debate que evite ideas confusas o distorsionadas.

La idea no es únicamente generar debate, sino llegar a obtener información, lo más completa posible, sobre el acontecimiento estudiado.

Las fuentes históricas son cualquier testimonio (escrito, oral, material) que permite la reconstrucción, el análisis y la interpretación de los acontecimientos históricos. Las fuentes históricas constituyen la materia prima de la Historia.

La diversidad de fuentes históricas puede ser objeto de diferentes clasificaciones según su origen, el soporte en el que se encuentran, la temática que abordan o a la que se refieren, la intencionalidad (si la tienen), etc.

Por su origen, las fuentes históricas se clasifican en fuentes primarias o directas y fuentes secundarias, indirectas o historiográficas.

Las FUENTES PRIMARIAS proceden de la época que se está investigando. Son testimonios de primera mano contemporáneos a los hechos: leyes, tratados, memorias, censos de población, artículos de prensa, imágenes, objetos de la vida cotidiana...

Las FUENTES SECUNDARIAS han sido elaboradas con posterioridad al período que se está estudiando y son obra de los historiadores. Fundamentalmente son los libros de texto, los manuales, los estudios científicos y artículos de revistas especializadas... También son fuentes secundarias los gráficos y los mapas temáticos realizados con datos primarios.

Según el soporte en el que se presentan, las fuentes históricas pueden clasificarse en:

Fuentes ESCRITAS o TEXTUALES: Son las fuentes más habituales y pueden ser primarias o secundarias. Las primarias son los documentos jurídicos (leyes i testamentos) y textos oficiales, las memorias, crónicas, censos y registros parroquiales, cartas, diarios privados, prensa y ensayos de la época, textos literarios del momento... Las secundarias son los libros de historia y otros trabajos de los historiadores.

Fuentes GRÁFICAS Y ESTADÍSTICAS: Normalmente son fuentes secundarias en las cuales se muestran datos numéricos sobre determinados temas de carácter económico, demográfico, climático, etc. Son fuentes de información cuantitativa y se representan habitualmente en forma de tabla de datos o gráficos.

Fuentes ICONOGRÁFICAS: Son fuentes primarias e incluyen cualquier tipo de imágenes: la pintura, las fotografías, los grabados, les ilustraciones y los carteles, las caricaturas, los cómics, etc.

Fuentes CARTOGRÁFICAS: Son los mapas. Raramente son fuentes primarias. Los mapas tienen un lenguaje específico que es necesario poder interpretar, para lo que debe adjuntarse una clave (o leyenda) con el significado de los símbolos, los colores o las tramas utilizados en su realización. Los mapas históricos son temáticos y de dos tipos: sincrónicos (o estáticos) (explican la situación en un momento determinado) y diacrónicos (o dinámicos) (explican la evolución de una situación histórica y los cambios que se han producido).

Fuentes MATERIALES (restos materiales y construcciones, objetos personales, herramientas, monedas, armas, objetos decorativos, etc): Son fuentes primarias y proporcionan información sobre diversos aspectos, como el poder, la riqueza, la sociedad, la vida cotidiana y las costumbres, los gustos y las modas.

Fuentes ORALES: Pueden ser testimonios directos o grabaciones en diferentes soportes. La entrevista es la fuente más habitual, pero también se incluyen los discursos, los programas de radio, las canciones, los cuentos... Son fuentes primarias.

• Según su temática, teniendo en cuenta que las fuentes históricas tratan y nos dan información sobre alguno o diversos aspectos (política, economía, sociedad, ciencia y técnica, religión, cultura, arte), pueden ser: fuentes políticas, económicas, técnicas y científicas, sociales, religiosas, militares... Las leyes, en cualquiera de sus formas (pragmática, decreto, constitución...) son fuentes jurídicas o legislativas.

• Finalmente, según su intencionalidad, aunque la mayor parte de las fuentes históricas son exclusivamente informativas, en algunas, como las caricaturas, discursos políticos, o algunos artículos de prensa, se observa una clara intencionalidad crítica, satírica o panegírica (de elogio a alguien) que si resulta muy evidente es conveniente destacar. Por su parte, las fuentes jurídicas o legislativas tienen una intencionalidad (y función) normativa y reguladora.

4.1.4 Fenómenos Demográficos.

¿Qué son los fenómenos demográficos?

Se trata de tendencias o indicadores que provienen de la información demográfica; estudios estadísticos de las poblaciones humanas según su estado y distribución en un contexto particular, ya sea su posición geográfica, su evolución histórica o su contexto social.

¿Qué son las poblaciones?

Son grupos que permanecen el suficiente tiempo en un mismo lugar para ser identificables y están constituidas de individuos que forman vínculos y pueden distinguirse por características religiosas, políticas, jurídicas, étnicas o territoriales.

Las poblaciones se extinguen, mezclan y cambian; esto puede explicarse a partir de los fenómenos demográficos como lo es la natalidad, la mortandad y la movilidad, además de la migración, la inmigración, la urbanización y el despoblamiento.

Las combinaciones entre estos fenómenos demográficos determinan la variación numérica de una población y, de esta manera, se puede comprender por qué se encuentran grupos en fuerte crecimiento y otros en decaimiento. Éstos también indican factores como la cohesión y la repulsión.

Las condiciones biológicas, económicas, culturales e históricas afectan e influyen considerablemente a los fenómenos demográficos, acentuándolos o debilitándolos. Por eso son referencia para estudios de otra índole como la antropología, la sociología, la comunicación, entre muchos otros.

¿Cómo obtiene su información?

La forma en que los fenómenos demográficos acopian la información es a partir de censos; desde un conocimiento general sobre la población hasta sus características particulares como edad, género, escolaridad, etcétera.

La forma en que se recaba esta información es a partir de unidades generalmente conformadas por familias: un grupo de personas de dos o más individuos que residen en la misma vivienda, comparten gastos y lazos de parentesco, ya sea de sangre o políticos.

Un proceso tan largo y complejo evidentemente tiene cierto margen de error; la doble contabilización y la omisión son los problemas más comunes. Sin embargo, existen métodos para promediarlo y llegar a estadísticas válidas.

Los fenómenos demográficos son importantes porque son un esbozo de la realidad que vivimos a nivel mundial: población, natalidad, mortandad, migración e inmigración. Gracias a ellos podemos proponer políticas que resuelvan crisis sociales, mejorar la calidad de vida y prevenir problemas futuros.

Susana Torrado define a los fenómenos demográficos (nupcialidad, fecundidad, etc.) en base a una serie de características entre las que se destaca que los fenómenos demográficos (en adelante: FD) constituyen un sistema ya que están mutuamente interrelacionados y cualquier variación en las tendencias de uno repercute en las de los demás. Este sistema ubica a la población en un espacio geográfico dando lugar a ese sistema; pero a su vez en un mismo espacio los FD se pueden comportar de forma diferencial según los estratos sociales y las regiones geográficas. En otras palabras, el comportamiento de los hogares esta interrelacionado con la población que conforman influenciando y siendo influenciada por el entorno. "La dinámica demográfica resume los hechos relacionados con el tamaño, la composición y la distribución espacial de la población, los cuales dependen de los siguientes fenómenos de población: nupcialidad, fecundidad, mortalidad y migraciones (internas e internacionales), es decir, son de causación recíproca y diferenciales por clase social y por región geográfica".

Los datos demográficos surgen de una agregación sistémica de los hogares que son la unidad mínima de reproducción de la población en tanto fuerza de trabajo y su principal aplicación tiene que ver con poder describir el contexto de agrupamiento en que se dan ciertas dinámicas. En un sistema capitalista, es esperable que la mayor parte de los ingresos necesarios para la reproducción de la existencia provenga del trabajo de aquellos miembros del hogar en edad activa (15-64) mientras que se benefician de ello, aquellos en edad inactiva (menores hasta 14 y mayores de 64). Este principio, que registra muchas salvedades en la aplicación cotidiana, sirve como guía para ver la relación entre las dinámicas de generación y distribución de la renta y las dinámicas poblacionales. Por ello, la primera coordenada de análisis apunta a ver la distribución etérea de la población en términos de aquellos que potencialmente pueden participar en el mercado de trabajo y generarse ingresos y aquellos que dependen de los primeros. Sin embargo, debe destacarse que esta dinámica depende de la estrategia de desarrollo nacional que regula las formas en que se producen y reproducen las condiciones de vida de acuerdo a las diferentes clases sociales, a las relaciones de fuerza entre ellas, y a las características históricas específicas de la clase dominante.

En este sentido, es importante considerar no sólo la composición de la población en términos de potencialmente activa/inactiva, sino además en qué medida estas estrategias de desarrollo contribuyen al mejoramiento o empeoramiento de las condiciones de vida de los hogares. Una segunda coordenada de interés en la distribución espacial de la población, es definida no sólo en términos de concentración sino de desigualdad territorial. Para una buena parte de la literatura, el término territorio ya de por sí hace referencia a desigualdad por cuanto circunscribe dinámicas de estratificación y diferenciación social que pueden volverse incluso conflictivas.

La demografía ha estudiado la concentración geográfica destacando que es un hecho relevante de los fenómenos demográficos. De hecho, uno de los rasgos más relevantes del proceso de modernización es la concentración de la población en grandes aglomerados urbanos. Esta concentración tiene como efecto secundario una agudización de ciertas dinámicas de desigualdad. Recientemente, Gabriel Kessler sostiene que, en la última década, la concentración de la población en aglomerados potenció la desigualdad existente en materia de infraestructura y de transporte entre zonas con distintos grados de integración a los núcleos urbanos más dinámicos de la economía.

Bibliografía básica y complementaria:

- Título del libro: BIOESTADISTICA WAYNE W. DANIELLIMUSA SA DE CV 2011.
- Título del libro: BIOESTADISTICA ALFREDO DE JESUS CELIS DE LA ROSA EL MANUAL MODERNO 2008.