



NOVEDADES EN POBLACIÓN

REVISTA ESPECIALIZADA EN TEMAS DE POBLACIÓN

Año 1

Número 1

enero-junio 2005

**“DIFUSIÓN DE LOS RESULTADOS CIENTÍFICOS:
DE LA NECESIDAD A LA REALIDAD”**

NOVEDADES EN POBLACIÓN

DIRECTORA:

Dra. Otilia Barros Díaz

DIRECTORA EJECUTIVA:

Dra. María Elena Benítez Pérez

EDITORIA JEFA:

Lic. Yolanda Morejón Bravo

EDICIÓN Y DISEÑO:

Irma Muñoz Viqueles

WEBMASTER:

Lic. Alejandro Sánchez Ravelo

CONSEJO ASESOR DEL NÚMERO:

Dra. María Elena Benítez Pérez

Dra. Sonia I. Casasús Cervera

Dr. Raúl Hernández Castellón

Dra. Norma Montes Rodríguez

REDACCIÓN:

*Centro de Estudios Demográficos
de la Universidad de La Habana
Avenida 41 #2003 e 20 y 22, Playa,
Ciudad de La Habana.*

Telefs: 202 81 41

202 81 85

202 82 61

Fax: (537) 204 06 30

Email: biblioteca@cedem.uh.cu

*Publicación semestral especializada en temas de
población. Los artículos publicados en esta revista
poseen la aprobación del Consejo Científico del
CEDEM y son un reflejo del punto de vista de los
autores.*

ISSN: en tramitación

CARTA DE LA DIRECTORA:

El Centro de Estudios Demográficos (CEDEM) es una institución de investigación, formación, capacitación y difusión de todo aquello relacionado con la población. Perteneció a la Universidad de La Habana, Cuba. Fue creado el 9 de febrero de 1972, y desde sus inicios tiene entre sus objetivos fundamentales la enseñanza académica de la demografía y las interrelaciones entre población y desarrollo, la investigación sobre la población como categoría demográficamente identificable y la información científica relacionada con estas temáticas.

El CEDEM se ha consolidado como un centro de excelencia académica e investigativa, lo que ha sido reconocido por la comunidad científica tanto nacional como internacional. Cuenta con un claustro de profesores e investigadores de alto nivel científico, de los cuales, el 87,5 % ha alcanzado el grado de Doctor o Master. Dispone del uso de las nuevas tecnologías en su equipamiento y de una biblioteca especializada con bibliografía actualizada en temas demográficos.

La difusión de los resultados de investigación resulta siempre de gran importancia porque cierra el círculo deseado para cualquier trabajo. Por eso, nos complace celebrar el aniversario XXIII de la creación del CEDEM, con la publicación del primer número de la *Revista Especializada en Temas de Población*. Los invitamos a su lectura. Confiamos sirva de ejemplo de cómo trabajamos en CEDEM.

Nos interesa mucho su opinión y los invitamos a publicar con nosotros. Contactenos:

**CENTRO DE ESTUDIOS DEMOGRAFICOS
UNIVERSIDAD DE LA HABANA**

**Avenida # 41 N° 2003, esquina a 20.
Municipio Playa. Ciudad de La Habana, Cuba**

**Teléfonos: (537) 202-8141, 202-8185, 202-8261
E-mail: biblioteca@cedem.uh.cu**

La Estadística como factor del desarrollo

Autora: Cristina López-Calleja Hiort-Lorenzen
Centro de Estudios Demográficos
Universidad de La Habana
E-mail: cristina@cedem.uh.cu

Resumen

En este artículo se trata de la importancia de la estadística matemática en las investigaciones sociales y de la imprescindible comunicación entre los especialistas de las distintas ramas del saber con los estadísticos. Se ofrecen algunos elementos conceptuales de la estadística, con sencillos ejemplos y se incluye una somera explicación sobre el método estadístico de contraste de hipótesis.

Introducción

Para una comprensión intuitiva y real de lo que es la Estadística como factor de desarrollo, es necesario en primer lugar que los profesionales de las distintas disciplinas puedan entenderse con el profesional de la Estadística Matemática, mediante un mismo lenguaje que los comunique, y como premisa se hace necesario que se comprenda que al igual que el estadístico no puede sustituir a otro profesional en cuanto a los conocimientos que tiene sobre su disciplina, éstos no pueden sustituir al estadístico, el cual requiere para su trabajo, del conocimiento de la teoría de Probabilidades, base teórica de la Estadística, y la cual se basa en un profundo conocimiento del Análisis Matemático, del Álgebra, de la Teoría de la Medida, de la Teoría de los Procesos Estocásticos, etc.

Pretendemos tratar de explicar en este artículo que el trabajo de investigación debe ser mancomunado y tiene numerosas etapas. Desde el momento en que el especialista de una determinada disciplina comienza a concebir la idea de una posible investigación, que responda a un problema planteado, y tenga claridad en cuanto a los objetivos que se propone alcanzar, este especialista debe consultar al estadístico matemático, que es el especialista que tiene el conocimiento sobre los métodos estadísticos que se pueden aplicar para lograr esos objetivos a alcanzar, o puede indagar sobre éstos.

Actualmente existen diferentes métodos para llegar a los objetivos planteados y que difieren entre sí, en primer lugar, en dependencia del conocimiento que tenga el especialista sobre el tema a investigar.

Si se conoce poco del fenómeno a investigar, es necesario hacer **análisis exploratorios univariados y multivariados** que pueden consistir en aplicar métodos descriptivos, como tablas cruzadas de frecuencias, gráficos de diferentes tipos, que incluyen gráficos de correspondencias, etc., análisis de componentes principales y de correspondencias, de agrupamiento (cluster análisis), etc., los cuales pueden dar una imagen que sea capaz de mostrar lo que pasa con los datos o información que se tenga del problema. En especial los fenómenos sociales se caracterizan por la interacción de diferentes variables que actúan simultáneamente en ellos, y son precisamente los métodos estadísticos multivariados los que permiten analizar los efectos inter actuantes y simultáneos de dichas variables

Una vez que se conoce cómo y de donde se va a tomar la información que requiere el trabajo de investigación, y se ha logrado obtener dicha información, es necesario elaborar la correspondiente base de datos, y para esto existen técnicas y programas estadísticos computarizados especialmente diseñados para ello, que se deben conocer y utilizar. Una base de datos deficiente implica un trabajo sin calidad.

En el análisis de la información mediante una base de datos que tenga la imprescindible calidad, los gráficos que se elaboran tienen una gran importancia.

El aserto "una imagen vale más que mil palabras" se puede aplicar al ámbito de la estadística descriptiva diciendo que "un gráfico bien elaborado vale más que mil tablas de frecuencias". Cada vez es más habitual el uso de gráficos o imágenes para representar la información obtenida. No obstante, debemos ser prudentes al confeccionar o interpretar gráficos, puesto que una misma información se puede representar de formas muy diversas, y no todas ellas son pertinentes, correctas o válidas.

Si queremos hacer un uso de la Estadística como factor del desarrollo, debemos tener una clara idea sobre,

¿Qué es la estadística?

Cuando coloquialmente se habla de estadística, se suele pensar en una relación de datos numéricos presentada de forma ordenada y sistemática. Esta idea es la consecuencia del concepto popular que existe sobre el término y que cada vez está más extendido debido a la influencia de nuestro entorno, ya que hoy día es casi imposible que cualquier medio de difusión, periódico, radio, televisión, etc. no nos aborde diariamente con cualquier tipo de información estadística sobre accidentes de tráfico, índices de crecimiento de población, turismo, tendencias políticas, calidad de vida, pobreza, etc.

Sólo cuando nos adentramos en un mundo más específico como es el campo de la investigación de las Ciencias Sociales: Sociología, Psicología, Economía, Demografía, Medicina... empezamos a percibir que la Estadística no sólo es algo más, sino que se convierte en la herramienta que, hoy por hoy, permite dar luz y obtener resultados, fundamentados rigurosamente, y por tanto obtener beneficios en cualquier tipo de estudio, cuyos movimientos y relaciones, por su variabilidad intrínseca, no puedan ser abordadas desde la perspectiva de las leyes deterministas.

George Box, eminente estadístico dijo “ *La Estadística es, con mucho, demasiado importante como para dejarla por completo a los estadísticos*”. Poco después, Walt Federer sostuvo. “ *La ciencia es, con mucho, demasiado importante como para dejarla por completo a los científicos*”. ¡Estos dos famosos estadísticos estaban en lo correcto! Nunca antes en la historia de la ciencia y la estadística ha existido una mayor necesidad para que haya interacción y colaboración entre los científicos de distintas ramas del saber y los matemáticos y estadísticos.

Durante los últimos 20 años se han impartido cursos y seminarios universitarios sobre análisis estadísticos multivariados aplicados. Muchas preguntas importantes pueden ayudarse a responder mediante métodos multivariados. A medida que crece el tamaño de los conjuntos de datos, los métodos multivariados se vuelven más útiles.

Las actuales tecnologías facilitan la recolección de grandes cantidades de datos; se necesitan los métodos multivariados para determinar si, en realidad, esas cantidades masivas de datos contienen información. Se ha dicho que aunque no es fácil reunir datos, es mucho más difícil reunir información. Los métodos estadísticos multivariados pueden ayudar a determinar si existe información en los datos y también pueden ayudar a resumir esa información.

Definitivamente *los métodos estadísticos multivariados son demasiado importantes como para que únicamente se enseñen a los estadísticos*. Es muy necesario que los especialistas de distintas ramas del saber puedan responder a las siguientes preguntas:

- ¿Por qué se deben usar los métodos estadísticos multivariados?
- ¿Cuándo se deben usar?
- ¿Cómo se pueden usar?
- ¿Qué se ha aprendido por la aplicación de estos métodos?

Es usual escuchar a científicos sociales expresar que su disciplina no admite sólo análisis cuantitativos, como creen que hacen los estadísticos, pues los análisis cualitativos son los que imperan en sus disciplinas, y por esa errática idea, rompen con un posible análisis estadístico.

¿Cómo es posible que personas con profundos conocimientos de su disciplina desconozcan de tal manera, lo que la Estadística puede brindarle? ¿Es que piensan que los estadísticos sólo realizan análisis cuantitativos, y no tienen nada que ver con la calidad y las cualidades de un fenómeno?

Esos criterios que se basan en una gran falta de comunicación y por ende ignorancia sobre la Estadística, les lleva a que pierdan la posibilidad de intercambiar criterios con los estadísticos y así estructurar equipos multidisciplinarios para el trabajo científico investigativo, como cada día más se hace en los centros y universidades, que han alcanzado un gran desarrollo científico.

En este artículo se hará un intento de explicar en qué consiste la Estadística y la necesidad de su uso, y aspiramos a que se entienda someramente uno sólo de los métodos generales de la Estadística mediante un sencillo ejemplo.

Podríamos, desde un punto de vista más amplio, definir *la Estadística* como la *ciencia que estudia cómo debe emplearse la información y cómo dar una guía de acción en situaciones prácticas que entrañan incertidumbre*.

La Estadística se ocupa de los métodos y procedimientos para recoger, clasificar, resumir, hallar regularidades y analizar los *datos*, siempre y cuando la variabilidad e

incertidumbre sea una causa intrínseca de los mismos; así como de realizar *inferencias* a partir de ellos, con la finalidad de ayudar a la toma de *decisiones* y en su caso formular *predicciones*.

Podríamos por tanto clasificar la Estadística en descriptiva, cuando los resultados del análisis no pretenden ir más allá del conjunto de datos, e inferencial cuando el objetivo del estudio es derivar las conclusiones obtenidas a un conjunto de datos más amplio.

Estadística descriptiva: Describe, analiza y representa un grupo de datos utilizando métodos numéricos y gráficos que resumen y presentan la información contenida en ellos.

Estadística inferencial: Apoyándose en el cálculo de probabilidades y a partir de datos muestrales, se efectúan estimaciones, decisiones, predicciones u otras generalizaciones sobre un conjunto mayor de datos.

Estadística bayesiana: Mediante las denominadas probabilidades de Bayes se obtienen estimaciones a partir de los datos muestrales.

Elementos. Población. Caracteres

Establecemos a continuación algunas definiciones de conceptos básicos como son: elemento, población, muestra, caracteres, variables, etc., que son parte del vocabulario básico de los estadísticos.

Individuos o elementos: personas u objetos que contienen cierta información que se desea estudiar

Población: conjunto de individuos o elementos que cumplen ciertas propiedades comunes.

Muestra: subconjunto que representa a una población.

Parámetro: función definida sobre los valores numéricos de características medibles de una población.

Estadígrafo: función definida sobre los valores numéricos de una muestra.

Con relación al tamaño de la población, ésta puede ser:

- **Finita**, como es el caso del número de personas que llegan al servicio de urgencia de un hospital en un día;
- **Infinita**, si por ejemplo estudiamos el mecanismo aleatorio que describe la secuencia de caras y cruces obtenida en el lanzamiento repetido de una moneda al aire.

Un Ejemplo bastante conocido de estadígrafo

Consideremos la población formada por todos los estudiantes de la Universidad de la Habana (finita). La *altura media de todos los estudiantes* constituye el *parámetro* μ . El conjunto formado por una selección proporcional a la cantidad de alumnos de cada una de las facultades de dicha universidad es una *muestra* de dicha población y la altura media de esta muestra es un *estadígrafo* que se denomina *media muestral* y se anota por \bar{X} , el valor que toma es *una estimación de* μ .

Media muestral

Si se anotan por $x_1, x_2, x_3, \dots, x_n$ las alturas de los n estudiantes de la muestra, se tiene que:

$$\bar{X} = \frac{x_1 + x_2 + x_3 + \dots + x_n}{n} \quad \text{ó} \quad \bar{X} = \frac{\sum_{i=1}^n x_i}{n} .$$

Se dice que, en este ejemplo, la altura es la *variable aleatoria* X que toma los valores $x_1, x_2, x_3, \dots, x_n$, y el valor estimado del parámetro μ lo proporciona el estadígrafo \bar{X} .

Las variables aleatorias miden los caracteres o características: propiedades, rasgos o cualidades de los elementos de la población. Estos caracteres pueden dividirse en *cualitativos* y *cuantitativos* y corresponden a las *variables discretas* y *continuas*.

Categorías de una variable discreta: diferentes situaciones posibles de una característica. Las categorías deben ser a la vez exhaustivas y mutuamente excluyentes, es decir, cada elemento posee una y sólo una de las categorías posibles.

Escalas de medición o cuantificación: Una medición consiste en establecer una correspondencia entre un conjunto de manifestaciones de una propiedad y un conjunto de entes que se asumen como los valores de la medición de dicho conjunto. Una vez que se concreta el proceso de medición, se puede determinar qué procesamiento estadístico puede realizarse.

Existen dos tipos básicos de datos: *los métricos (cuantitativos)* y *los no métricos (cualitativos)*. Los datos de *medidas no métricos* son atributos, características o propiedades categóricas que identifican o describen a un sujeto; Describen diferencias en tipo o clase, indicando la presencia o ausencia de una característica o propiedad. Muchas propiedades son discretas porque tienen una característica peculiar que excluye a todas las demás características. Por ejemplo, si una persona es hombre, no puede ser mujer. No hay cantidad de “genero”, sólo la condición de ser hombre o mujer.

Por el contrario, las *medidas de datos métricos* son de tal forma que los sujetos pueden ser identificados por diferencias entre grado o cantidad, reflejan cantidades relativas o grado. Las medidas métricas son las más apropiadas para casos que involucren cantidad o magnitud, tales como el nivel de satisfacción o la demanda de trabajo.

Se pueden considerar cuatro escalas de medición: *nominal, ordinal, de intervalo y de razón o proporción*, y en dependencia de ellas se aplicarán los métodos específicos de la Estadística.

Al medir con una *escala nominal* se asignan números que se usan para identificar objetos o sujetos. Las escalas nominales se conocen también como escalas de categorías, los números o símbolos asignados no tienen más significado cuantitativo que indicar la presencia o la ausencia del atributo o la característica bajo investigación. Son datos con escala nominal, el sexo, la religión, o la integración política de una persona. El analista puede asignar el número 1 al hombre y el número 2 a la mujer, etc.

La *escala ordinal* representa un nivel superior de precisión de la medida. Las variables pueden ser ordenadas o clasificadas con escalas ordinales en relación a la cantidad del atributo que posee el sujeto u objeto. Por ejemplo, los diferentes niveles de satisfacción del consumidor individual por diferentes productos se pueden medir por una escala ordinal. El nivel de satisfacción de un encuestado por tres productos, A, B y C.



Los números utilizados en escalas nominales no son cuantitativos, dado que indican sólo posiciones relativas en series ordenadas. En este ejemplo, no hay medida de cuanta satisfacción recibe el encuestado en términos absolutos, el investigador ni siquiera conoce la diferencia exacta entre puntos de la escala de satisfacción. Muchas escalas de las ciencias del comportamiento corresponden a una escala ordinal.

Las *escalas de intervalo y de razón* para datos de medida métricos proporcionan el nivel más alto de precisión en la medida. Estas dos escalas tienen unidades constantes de medida, de tal forma que las diferencias entre dos puntos adyacentes de cualquier parte de la escala son iguales. La única diferencia entre ambas escalas es que la de intervalo tiene un punto que es un cero arbitrario, mientras que la de razón tiene un punto que es un cero absoluto.

Se miden en escala de intervalo la temperatura, en grados Celsius o en grados Fahrenheit. Ambas temperaturas tienen un cero arbitrario, pero ese cero no indica una cantidad cero o ausencia de temperatura, ya que se pueden registrar temperaturas por debajo del punto cero de esa escala. No es posible decir que un valor situado en un punto de la escala es un múltiplo de cualquier otro punto de la escala. Por ejemplo si en un día se registran 80° F , no se puede decir que ese día sea dos veces más caluroso que otro en el que se registraron 40° F, aunque $80 = 2 (40)$, ya que se sabe que 80° F equivale a 26,7° C y 40° F equivale a 4,4° C y al utilizar diferentes escalas, se tiene que el calor no es dos veces mayor, esto es, $26,7^{\circ} C \neq 2 (4,4^{\circ} C)$.

La *escala de razón* representa la forma superior de medida, con ella se permiten todas las operaciones matemáticas, es la única que tiene un punto de cero absoluto y una medida en esta escala de razón puede ser expresada en términos de múltiplo de otra cualquier medida de la escala.

El peso de una persona se mide en una escala de razón y por ejemplo, se puede decir que 100 kg es dos veces más pesado que 50 kg.

Es importante que el investigador identifique bien la escala en que mide las variables para poder determinar que método estadístico puede utilizar.

Otro ejemplo: En el estudio de un grupo de niños discapacitados:

VARIABLE	ESCALA
El tipo de discapacidad	Nominal
El grado de la discapacidad	Ordinal
La temperatura corporal en grados centígrados	De intervalo
El peso en kilogramos y la edad en años	De razón

Varianza Muestral

De la misma forma se utiliza el estadígrafo **varianza muestral** que no es más que la media aritmética de los cuadrados de las diferencias de los valores x_i y la media muestral, es decir,

$$S^2 = \frac{(x_1 - \bar{X})^2 + (x_2 - \bar{X})^2 + (x_3 - \bar{X})^2 \dots + (x_n - \bar{X})^2}{n} \quad \text{ó} \quad S^2 = \frac{\sum_{i=1}^n (x_i - \bar{X})^2}{n}$$

y su raíz cuadrada que se anota por **S** es la **desviación típica o estandar de la variable X**, que también es un estadígrafo que se usa más que la varianza muestral, debido a que se mide en las mismas unidades que la media muestral de la variable **X**.

$$S = \sqrt{\frac{(x_1 - \bar{X})^2 + (x_2 - \bar{X})^2 + (x_3 - \bar{X})^2 \dots + (x_n - \bar{X})^2}{n}} \quad \text{ó} \quad S = \sqrt{\frac{\sum_{i=1}^n (x_i - \bar{X})^2}{n}}$$

Moda

Otro estadígrafo muy utilizado es **la moda** que es el valor o dato más frecuente de la variable, es decir el dato que más se repite.

En este ejemplo la moda es 30 años.

Mediana

El estadígrafo llamado **mediana** es el elemento que ocupa la posición central en la lista de los datos ordenados de menor a mayor. La mediana es la anotación que supera al 50 % de los datos y es superada por el 50 % de los datos.

En distribuciones de variables discretas no tiene que coincidir con un dato.

En las siguientes muestras la mediana es:

Muestra	Mediana	
a) 1, 2, 5, 7, 9	5	
b) 1, 2, 5, 7	3,5	pues $\frac{2+5}{2} = 3,5$

Tabla de contingencia o tabla del cruce de dos variables o tabla de doble entrada (Crosstab en inglés)

Una tabla de contingencia de las variables X y Y se puede anotar como sigue:

X\Y	Y ₁	Y _j	Y _J	
X ₁	n ₁₁	n _{1j}	n _{1J}	n _{1.}
:	:	:	:
X _i	n _{i1}	n _{ij}	n _{iJ}	n _{i.}
:	:	:	:
X _I	n _{I1}	n _{Ij}	n _{IJ}	n _{J.}
	n _{.1}	n _{.j}	n _{.J}	n

Siendo n_{ij} la frecuencia de individuos asociados a la categoría i de la variable X y a la categoría j de la variable Y simultáneamente, se dice que n_{ij} es la **frecuencia de la celda o casilla i j** y no es más que el número de veces que el par (X_i, Y_j) aparece en las

observaciones, donde $i = 1, 2, \dots, I$; y $j = 1, 2, \dots, J$. Además $n_{i \cdot}$ es la frecuencia del valor X_i de la variable X y $n_{\cdot j}$ es la frecuencia del valor Y_j de la variable Y y $n = \sum_{i,j} n_{ij}$.

Correlación entre variables:

Coefficiente de correlación lineal de Pearson:

$R_{X,Y} = \frac{S_{X,Y}}{S_X \cdot S_Y}$ siendo $S_{X,Y} = \frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})$ la **covarianza** entre las variables X y Y .

También se cumple que la **covarianza** es: $S_{X,Y} = \sum_{i=1}^n x_i \cdot y_i \cdot \left(\frac{n_{ij}}{n} \right) - \bar{x} \cdot \bar{y}$

COEFICIENTE DE CORRELACIÓN	ESCALA QUE EXIGE
Lineal de Pearson	de intervalo
De rangos de Spearman	ordinal
De rangos de Kendall	ordinal
Punto biserial	una dicotómica y otra de intervalo
V de Cramer	Nominal
ϕ (caso particular de V)	nominal dicotómica

El coeficiente de correlación de Pearson sólo puede tomar valores en el intervalo $[-1; 1]$, si $R_{X,Y} = 1$ entonces todas las observaciones muestrales (X_i, Y_j) pertenecen a la representación gráfica de una recta en el plano cartesiano, y se dice que la correlación lineal entre X y Y es positiva y perfecta, si $R_{X,Y} = -1$ es negativa la correlación es perfectamente negativa. Si $R_{X,Y} = 0$ no hay correlación lineal. Si la $R_{X,Y}$ se aproxima a cualquiera de los valores $-1, 0$ ó 1 , la correlación será aproximadamente negativa, ausente o positiva respectivamente.

Existen muchos métodos estadísticos para el estudio de muy diferentes problemas, pero es necesario que se tenga en cuenta que la metodología estadística se concibe como un proceso iterativo de aprendizaje en lugar de concebirse como una aplicación única y directa de un determinado proceso óptimo.

Se analizará brevemente como ejemplo, *un método general* muy importante que brinda la Estadística, y que se aplica en muchos análisis estadísticos.

Pueden presentarse en la práctica, situaciones en las que exista una teoría preconcebida relativa a la característica de la población sometida a estudio. Tal sería el caso, por ejemplo si pensamos que un tratamiento nuevo puede tener un porcentaje de mejoría mayor que otro estándar, o cuando nos planteamos si los niños de las distintas comunidades cubanas tienen la misma altura. Este tipo de circunstancias son las que nos llevan al estudio del *método estadístico inferencial* que se define bajo el título genérico de *Contraste de Hipótesis o Docimacia*. Este método implica, en cualquier investigación, la existencia de dos teorías o hipótesis implícitas, que denominaremos *hipótesis nula e hipótesis alternativa*, que de alguna manera reflejarán esa idea a priori que tenemos y que pretendemos contrastar con la "realidad". De la misma manera aparecen, implícitamente, diferentes *tipos de errores* que podemos cometer durante el procedimiento. No podemos olvidar que, habitualmente, el estudio y las conclusiones que obtengamos para una población cualquiera, se habrán apoyado exclusivamente en el análisis de sólo una parte de ésta. *De la probabilidad con la que estemos dispuestos a asumir estos errores, dependerá, por ejemplo, el tamaño de la muestra requerida y del método a utilizar*. Para comenzar se extrae una muestra aleatoria de dicha población.

Los *contrastos de significación o pruebas de hipótesis* se realizan: suponiendo *a priori* que la ley de distribución de la población es conocida.

Si la distribución de la muestra es "diferente" de la distribución de probabilidad que hemos asignado *a priori* a la población, concluimos que probablemente sea errónea la suposición inicial.

Ejemplo:

Supongamos que debemos realizar un estudio sobre la altura media de los habitantes de cierto pueblo de Cuba. Antes de tomar una muestra, lo lógico es hacer la siguiente suposición a priori o hipótesis que se desea contrastar y que denotamos por H_0 : La altura media es $\mu = 1,65$.

Al obtener una muestra de tamaño $n = 8$, podríamos encontrarnos ante una de las siguientes muestras:

1ª. muestra = {1,50 ;1,52; 1,48; 1,55; 1,60; 1,49; 1,55; 1,63}

2ª. muestra = {1,65; 1,80; 1,73; 1,52; 1,75; 1,65; 1,75; 1,78}

Intuitivamente, en la muestra 1 sería lógico suponer que, salvo que la muestra obtenida de los habitantes del pueblo sea obtenida arbitrariamente, la hipótesis H_0 debe ser rechazada. En el caso de la muestra 2 tal vez no podamos afirmar con cierta seguridad que la hipótesis H_0 sea cierta, sin embargo no podríamos descartarla y la admitimos por una cuestión de simplicidad.

Este ejemplo sirve como introducción de los siguientes conceptos: En un *contraste de hipótesis* o también *prueba de hipótesis* o *contraste de significación*, se decide si cierta hipótesis H_0 que denominamos *hipótesis nula* puede ser rechazada o no a la vista de los datos suministrados por una muestra de la población. Para realizar el contraste se necesita establecer previamente la *hipótesis alternativa* (H_1) que será admitida cuando H_0 sea rechazada. Normalmente H_1 es la negación de H_0 , aunque esto no es necesariamente así.

El procedimiento general consiste en definir un estadígrafo T relacionado con la hipótesis que deseamos contrastar. A éste lo denominamos *estadígrafo del contraste* o *de la prueba de hipótesis*. A continuación suponiendo que H_0 es verdadera se calcula un intervalo denominado *intervalo de aceptación de la hipótesis nula*, de manera que al calcular el estadígrafo con los datos de la muestra observada $T = T_{\text{observado}}$ el criterio a seguir sea el siguiente:

El intervalo de aceptación o más precisamente, de no rechazo de la hipótesis nula se establece fijando una cantidad suficientemente pequeña denominada *nivel de significación* α , de modo que la probabilidad de que el estadígrafo del contraste tome un valor fuera de una región llamada *región crítica* o *de rechazo*, cuando la hipótesis nula es cierta sea inferior a α ; Esto se ha de entender como sigue:

Si H_0 es correcta el criterio de rechazo sólo se equivoca con probabilidad α que es la probabilidad de que una muestra dé un valor del estadígrafo del contraste fuera del intervalo de aceptación.

La decisión de rechazar o no la hipótesis nula está al fin y al cabo basado en la elección de una muestra tomada al azar, y por tanto es posible cometer decisiones erróneas. Los errores que se pueden cometer se clasifican como sigue:

Error de tipo I :

Es el **error que consiste en rechazar H_0 cuando es cierta**. La probabilidad de cometer este error es lo que anteriormente hemos denominado nivel de significación. Es una costumbre establecida el denotarlo siempre con la letra α .

Error de tipo II:

Es el error que consiste en **no rechazar H_0 cuando es falsa**. La probabilidad de cometer este error la denotamos con la letra β .

En la prueba de las hipótesis se tienen que controlar los dos tipos de errores α y β .

Esta explicación sobre el método general de Contraste de Hipótesis constituye una somera idea de una forma diferente de pensar y demostrar probabilísticamente si una hipótesis se puede considerar verdadera o no y cuál es el riesgo permisible de equivocarse.

Con el desarrollo de los medios de computación ha sido plausible la aplicación y el desarrollo teórico de variados métodos estadísticos, algunos de ellos ya han sido estructurados, y en un futuro inmediato y mediato surgirán muchos otros métodos para el análisis de nuevas y disímiles situaciones que se irán presentando y que requieren soluciones. Los métodos estadísticos incluyen el tratamiento de variables cualitativas, que han requerido de los avanzados medios de computación para su desarrollo teórico y práctico.

Por ejemplo: Los modelos bayesianos, los modelos Hazard y de Cox, los modelos multiniveles, etc.

Para lograr el conocimiento de todos estos nuevos modelos y teorías estadísticas, así como su aplicación, se requieren profundos conocimientos matemáticos y estadísticos, que sólo un estudioso o un especialista de la Estadística Matemática, puede lograr, siempre que posea los requeridos conocimientos de computación.

Hay que recordar que los métodos estadísticos no pueden probar que uno o varios factores tienen un efecto particular. Sólo proporcionan directrices para la veracidad y validez de los resultados.

Los métodos estadísticos aplicados adecuadamente, sólo hacen posible obtener el probable error de una conclusión o asignar un nivel de confiabilidad a los resultados. La principal ventaja de los métodos estadísticos es que agregan objetividad al proceso de toma de decisiones. Las técnicas estadísticas aunadas al conocimientos técnico o del proceso que se investiga y al sentido común, suelen llevar a conclusiones razonables.

Es por ello que se puede plantear que *la Estadística constituye un factor del desarrollo*, y éste se logrará mediante la interrelación estrecha de especialistas mancomunados en equipos multidisciplinarios, con la participación activa del estudioso de la Estadística Matemática, y su acierto dependerá de la comunicación y participación conjunta con los especialistas de la rama de la investigación de que se trate.

Bibliografía

- Jonson, Dallas E. , Métodos multivariados aplicados al análisis de datos, Ed. Internacional Thomson Editores S.A. de C.V. 2000, México.
- Freixas I. Blanxart, M., Salafranca I. Cosialls, L., y otros profesores de estadística del departamento de metodología de las Ciencias del Comportamiento de la universidad de Barcelona; Análisis exploratorios de datos: Nuevas técnicas estadísticas, Ed. Promociones y Publicaciones Universitarias S.A. (PPU), 1992, España.
- Montgomery, Douglas C., Diseño y Análisis de Experimentos, Ed. Grupo Editorial Iberoamérica S.A. de C.V., 1991, México.
- Análisis multivariante, Hair Jr., J.F., Anderson, R.F., Tatham, R.I., Black, W.C., Ed. Prentice Hall Iberia, 1999, Madrid, España.