



**unam - ents**

---

Universidad Nacional Autónoma de México Escuela Nacional de Trabajo Social

## **Estadística Aplicada a la Investigación Social II**

**Mtro. José Luis Sandoval Dávila**

Área: Metodología y  
Práctica de Trabajo Social

Semestre: 4

Créditos: 5

Carácter: Obligatoria

---

Sistema Universidad Abierta

## CONTENIDO

	Pág.
Presentación	3
Introducción	4
Objetivo	7
Perfil de egreso	8
Temario general	9
Unidades de estudio	
Unidad 1 Diseño de Muestras en la Investigación Social	12
Unidad 2 Hipótesis Estadísticas.	34
Unidad 3 Pruebas estadísticas de correlación e independencia entre variables	43
Glosario	62
Preguntas frecuentes	66
Bibliografía	67

## PRESENTACIÓN

La Escuela Nacional de Trabajo Social inició sus estudios de *Licenciatura en Sistema Universidad Abierta*, en el año escolar 2003, con el Plan de Estudios aprobado por el H. Consejo Universitario el 10 de julio de 1996. Fue reestructurado en el año 2002 con aprobación del Consejo Académico del Área de las Ciencias Sociales, en su sesión del 26 de noviembre de 2002.

En el Sistema Universidad Abierta, la relación entre asesores, estudiantes y material didáctico es fundamental. En este sentido, en la Escuela se puso especial atención para lograr mayor calidad en los materiales.

De ésta manera, el material que ahora te presentamos debe constituirse en una herramienta fundamental para tu aprendizaje independiente. Cada uno de los componentes que lo integran guardan una congruencia con el fin de que el estudiante pueda alcanzar los objetivos académicos de la asignatura.

El material pretende desarrollar al máximo los contenidos académicos, temas y subtemas que son considerados en el programa de estudio de la asignatura. Esto no pretende soslayar el papel y responsabilidad preponderante del estudiante, que debe profundizar en la búsqueda de conocimientos en todas aquellas fuentes que tenga a su alcance hasta hacer realidad los objetivos y el perfil de egreso propuesto.

Este material es perfectible, por ello, con el apoyo de las experiencias de los estudiantes y otros profesores, serán revisados y actualizados de manera permanente por el asesor. Cuyos aportes sin duda, contribuirán para su mejora y enriquecimiento.

Te damos la más cordial bienvenida y te deseamos toda clase de éxitos en los estudios que inicias en esta, tu Escuela: la **Escuela Nacional de Trabajo Social de la Universidad Nacional Autónoma de México.**

## INTRODUCCIÓN

En la investigación social es frecuente enfrentarse a la necesidad de disponer de información que formalmente evidencie la realidad de los fenómenos sociales, así como la detección y estudio de sus causas e impacto en sus efectos. Un factor determinante en ese proceso es la oportunidad con la que debe tenerse la información, ya que a partir de ella está sustentada la toma de decisiones. Otro de los factores es la naturaleza de las poblaciones en las que inciden esos fenómenos, en algunas de ellas son conocidos sus elementos y en otras no, de allí su clasificación en finitas e infinitas. Sin embargo, el investigador demanda los ingredientes que habrán de ser analizados como objeto de estudio, mismos que deberán ser obtenidos a partir de una elección: tomar una muestra de la población o realizar un censo de ella.

Ante esta disyuntiva debe considerarse el tipo de población, objetivo de la investigación y costo, factor que implica tiempo, recursos materiales, humanos y financieros; así como el impacto o costo social de la decisión que habrá de tomarse y que en muchos de los casos es el que determina las acciones a seguir.

Sin embargo, la naturaleza de las poblaciones puede simplificar el proceso, debido a que es posible observar su forma real a partir de tomar como referencia sólo una parte de ella, lo que se referirá como muestra poblacional. Pero qué confianza puede tenerse al observar sólo a una parte de la población y a partir de ella, emitir afirmaciones en relación con las características y diversos aspectos y condiciones en que se encuentra el universo en estudio. Cuántos elementos deben ser observados para ese propósito, de tal manera que posean la credibilidad necesaria para tomar las medidas conducentes al respecto.

La respuesta a esas preguntas es simple, ya que para estar cierto de 'las condiciones en que se encuentran los elementos de una población, sólo bastará con observar detalladamente una parte de ella y se llegará a la conclusión de que todos los individuos que la conforman tienen características

y actitudes comunes o similares, entre otras, que permiten su generalización a un menor costo; por supuesto que la observación detallada referida implica un proceso metodológico a seguir, lo que garantiza la efectividad de los resultados de la investigación.

Por tal virtud, se requiere que el investigador social posea un amplio dominio de la metodología para el diseño de muestras, donde cada una de las etapas que lo conforman resulta ser la más importante al momento de su aplicación. Soslayar este requisito es contribuir a que los investigadores recurran a diversas instancias para tal propósito, mismas que normalmente carecen del involucramiento necesario para lograr la efectividad de la investigación.

Como se ha señalado, cada una de las etapas del diseño de muestras es la más importante al momento en que se utiliza; sin embargo, la muestra no lo es todo en la investigación, pero sí las bases que la sustentan; por lo que no hay que perder de vista que una vez recogida la muestra le habrán de seguir un conjunto de métodos, técnicas, procedimientos para la obtención de la información requerida por la investigación. Surgirá una variedad de supuestos que tratarán de explicar el por qué del comportamiento de las variables analizando la relación entre ellas, conociendo sus rasgos importantes, su magnitud y diversos aspectos de interés tanto para la investigación como para el mismo investigador.

Estimado lector, el contenido de este material le ofrecerá diversos criterios que en la práctica profesional se emplean para el tratamiento y análisis estadístico de conjuntos de datos, considerando sobre todo que el profesional medio de las ciencias sociales no está plenamente identificado o familiarizado con los métodos estadísticos; por lo que se procura utilizar un lenguaje sencillo y simple que, sin deterioro de la calidad del contenido, permita la fácil comprensión de conceptos, términos y la realización de los ejercicios y casos prácticos que se incluyen, partiendo del supuesto de que los conceptos del primer curso de estadística han sido comprendidos.

Así mismo, posteriormente al proceso del diseño de la muestra, los datos serán analizados a través de la utilización de pruebas para la contrastación de hipótesis, mediciones estadísticas, pruebas de correlación entre variables de todo tipo.

Para el diseño y contrastación de hipótesis estadísticas, deberán considerarse diversos aspectos como los pasos que habrán de darse en su metodología, el tipo de prueba que se pretende realizar, las reglas para la decisión de aceptar o no los supuestos de la investigación, así como la búsqueda de explicación en el comportamiento de las variables, entre otros.

Las pruebas a realizar en este material, serán de correlación y de independencia entre variables, principalmente; considerando que en ciencias sociales, un gran número de investigadores utilizan variables con respuestas mutuamente excluyentes como "SI" y "NO", "estoy de acuerdo" y "no estoy de acuerdo" cuyas tendencias o comportamiento requieren de explicación estadística.

Se explicarán los pasos para la comprobación de hipótesis y se utilizarán estadísticos como la prueba **Jl** Cuadrada, **Q** de Kendall, **C** de Pearson y **r** de Pearson.

Básicamente, el curso de estadística inferencial está dividido en el proceso para el diseño de muestras, planteamiento y diseño de pruebas de hipótesis y la realización de pruebas estadísticas.

## **OBJETIVO GENERAL**

Conocer y aplicar la metodología para el diseño óptimo de muestras en una investigación social, así como la estructuración de hipótesis estadísticas y su contrastación.

## PERFIL DE EGRESO

Al finalizar el curso el alumno tendrá la habilidad para:

- Determinar las características para el diseño de muestras en una investigación social.
- Aplicar el proceso para la contrastación de hipótesis estadísticas.
- Identificar las características de las variables en estudio y estructurar las hipótesis estadísticas correspondientes.
- Utilizar el estadístico de prueba adecuado al tipo de variables en estudio y contrastar las hipótesis inherentes a la investigación.

Así mismo, adquirirá las aptitudes que le permitirán utilizar con efectividad los recursos estadísticos para la generación de información y la toma de decisiones en el contexto social principalmente.

## TEMARIO GENERAL

### I DISEÑO DE MUESTRAS EN LA INVESTIGACIÓN SOCIAL.

I.1. Conceptos básicos para el muestreo.

I.2. Etapas para el diseño de muestras.

I.2.1 Métodos de muestreo: Con reemplazo y Sin reemplazo.

I.2.2 Tipos de muestreo probabilística

I.2.2.1 Aleatorio.

I.2.2.2 Sistemático.

I.2.2.3 Estratificado

I.2.2.4 Polietápico.

I.2.3 Tipos de muestreo no probabilística

I.2.3.1 Por cuotas

I.2.3.2 Intencional.

I.2.3.3 Bola de nieve

I.2.3.4 Discrecional

I.2.4 Determinación del tamaño de la muestra

I.2.4.1 Nivel de confianza.

I.2.4.2 Margen de error.

I.2.4.3 Variabilidad.

## II HIPÓTESIS ESTADÍSTICAS

### II.1. Hipótesis Estadísticas

#### 11.1.1 Características de las hipótesis

### II.2. Hipótesis de trabajo

### II.3. Contraste de Hipótesis

### II.4. Hipótesis Nula

### II.5. Hipótesis Alternativa.

### II.6. Errores Estadísticos.

### II.7. Método para la comprobación de hipótesis estadísticas.

## III PRUEBAS ESTADÍSTICAS DE CORRELACIÓN E INDEPENDENCIA ENTRE VARIABLES.

### III.1. Conceptos.

### III.2. Prueba de correlación Q de Kendall.

### III.3. Prueba de correlación R de Pearson.

### III.4. Prueba de independencia entre variables Ji Cuadrada.

### III.5. Prueba de correlación C de Pearson.



## **UNIDAD 1**

### **DISEÑO DE MUESTRAS EN LA INVESTIGACIÓN SOCIAL.**

#### **INTRODUCCIÓN**

Empíricamente se ha demostrado que una de las principales áreas de oportunidad en la investigación social, es el diseño de la muestra; de allí la importancia de que el investigador conozca y domine las etapas para la selección de sus unidades de muestreo que serán objeto de estudio.

En esta unidad se presenta la metodología recomendada a seguir por los interesados en que sus investigaciones sean soportadas por un diseño óptimo de la muestra.

#### **OBJETIVO**

Conocer las etapas del proceso para el diseño óptimo de muestras, considerando forma y modo de la selección.

#### **TEMARIO**

##### **I. DISEÑO DE MUESTRAS EN LA INVESTIGACIÓN SOCIAL.**

I.1. Conceptos básicos para el muestreo

I.2. Etapas para el diseño de muestras.

I.2.1 Métodos de muestreo: Con reemplazo y Sin reemplazo

I.2.2 Tipos de muestreo probabilístico.

I.2.2.1 Aleatorio.

I.2.2.2 Sistemático.

I.2.2.3 Estratificado.

I.2.2.4 Polietápico.

I.2.3 Tipos de muestreo no probabilístico

I.2.3.1 Por cuotas. 1.2.3.2 Intencional.

I.2.3.3 Bola de nieve.

I.2.3.4 Discrecional.

I.2.4 Determinación del tamaño de la muestra.

I.2.4.1 Nivel de confianza.

I.2.4.2 Margen de error.

I.2.4.3 Variabilidad.

## **I.1. CONCEPTOS BÁSICOS PARA EL MUESTREO.**

¿Qué importancia o peso específico da usted a los conceptos? ¿Qué tan efectivo es un proyecto de investigación si se dominan los conceptos que se involucran en él? Para responder estas interrogantes, recuerde usted qué tan útil le ha resultado conocer o desconocer cómo funciona una computadora, o si sólo le interesa que cuando usted se sienta frente a ella, ésta encienda, le permita guardar o imprimir sus archivos o le sirva para enviar un mensaje por correo electrónico a algún amigo. Seguro que a usted sólo le interesa que haga las cosas; pasando a segundo plano todo el proceso que siguió el fabricante, desde el diseño hasta que se encontró en el escritorio de usted.

El objetivo de este material es que usted conozca de manera sencilla y práctica las aplicaciones de los métodos, técnicas y pruebas estadísticas utilizadas en una investigación social, recordándole que lo fácil se hace difícil por nosotros mismos, esto es sencillo.

**Población:** Es un conjunto de elementos bien determinados y con características comunes, que los hace formar parte del mismo. Ejemplo: mujeres solteras que trabajan en la industria farmacéutica en el turno matutino.

Si algún elemento del conjunto no contiene todas las características referidas, entonces no formará parte de él. Ejemplo: si alguna mujer trabajadora en esa industria no pertenece al turno matutino no formará parte de ese conjunto.

**Población finita:** Es aquel conjunto en el que el número de sus elementos es indefinido o indeterminado.  
Ejemplo: los empleados que recibieron crédito hipotecario el mes pasado.

**Población infinita:** Es aquel conjunto en el que el número de sus elementos es indefinido o indeterminado.  
Ejemplo: los trabajadores que recibirán préstamos para adquisición de automóvil durante los próximos diez años.

**Muestra:** Fracción o parte de la población. Ésta debe ser representativa del conjunto de donde ha sido extraída. Para ostentar el carácter de representatividad, la muestra deberá evidenciar tanto las coincidencias como las discrepancias existentes en la población.

**Parámetro:** Es todo valor o característica que ha sido calculado, observado o extraído de la población.

**Estimador:** También llamado estadístico, es todo valor o característica que ha sido calculado, observado o extraído de una muestra. El propósito del estimador es dar a conocer los parámetros sin observar a la totalidad de los elementos de la población.

Lo anterior sugiere que para determinar si un valor o característica es un parámetro o un estimador, deberá observarse su origen, si éste es la población será un parámetro; pero si surge de la muestra, entonces será un estimador.

Hacer estimaciones refiere que el investigador parte de conocer los resultados obtenidos de una muestra y referirlos confiablemente como los valores o características que posee la población. La estimación es la obtención de un valor o característica de la población, teniendo como origen una muestra. En el caso de las variables numéricas, la estimación puede ser puntual o por intervalos. Puntual es cuando se obtiene un sólo valor y por intervalos, cuando se conoce un rango en el cual se desplaza el parámetro o valor poblacional.

A partir del proceso del diseño de muestras, se pretende extender las conclusiones obtenidas a la totalidad de los elementos de la población, de allí la importancia y trascendencia de la metodología para la obtención de una muestra.

## **I.2. ETAPAS PARA EL DISEÑO DE MUESTRAS.**

La experiencia indica que una proporción significativa de los profesionales en ciencias sociales manifiesta cierta aversión por todo aquello relacionado con números, ya no se diga con matemáticas, como es el caso de la estadística y su amplia gama de recursos; sin embargo, resulta paradójico saber que para aplicar la estadística en las disciplinas sociales se requiere de ingredientes tan naturales en el ser humano como el sentido

común y la intuición, así como no perder de vista que las técnicas y los métodos son importantes, pero no determinantes, aunque usted no lo crea y le voy a explicar por qué.

Cuando se posee ese atributo que confiere a una persona la habilidad para analizar, la dificultad para el entendimiento, comprensión y operacionalización de la metodología estadística, se reduce notablemente, dando oportunidad para una mayor dedicación al análisis de las variables incidentes en el fenómeno estudiado y reduciendo consecuentemente el tiempo de operación, lo que garantizará información más oportuna y mejores decisiones.

Por tal virtud, debe enfatizarse en la importancia que el proceso de diseño de muestras tiene en el desarrollo de estudios e investigaciones y consistente en tres etapas que ofrecen respuesta a las preguntas ¿cómo? ¿quién? ¿cuántos? Es oportuno subrayar, que este proceso es más efectivo y garantiza mejores resultados, cuando es aplicable a poblaciones finitas.

### **I.2.1 MÉTODOS DE MUESTREO: CON REEMPLAZO Y SIN REEMPLAZO.**

Consiste en determinar cómo serán seleccionados los elementos que habrán de integrar la muestra y se clasifica en muestreo con reemplazo y sin reemplazo. El primero indica que los elementos de la población pueden pertenecer más de una vez a la muestra; quiere decir que si un individuo ya fue seleccionado, se le extraen los datos requeridos por la investigación y se integra nuevamente a su universo sin perder la oportunidad de volver a ser seleccionado para su estudio en la misma muestra. Por su parte, el muestreo sin reemplazo indica que si un elemento de la población ha sido extraído para la muestra, no volverá a tener la oportunidad de pertenecer nuevamente a ella.

Pregúntese y contéstese usted, en qué casos o cómo se determina el método de muestreo a seguir en un proceso de recolección de datos.

Una respuesta sencilla es que las características de la investigación determinan observar a un individuo en más de una ocasión y analizando las mismas variables, sólo que en tiempos diferentes.

Cuando en la metodología de una muestra no se especifica el método a seguir, considérese que es sin reemplazo.

### **I.2.2. TIPOS DE MUESTREO PROBABILÍSTICA.**

La etapa para determinar el tipo de muestreo a utilizar en una investigación o estudio reviste singular importancia, debido a que es a través de este procedimiento que se determina quién pasará a formar parte de la muestra, de tal manera que conserve la forma de su población, según lo comentado en párrafos anteriores.

El tipo de muestreo se clasifica en probabilístico y no probabilístico. El primero de ellos es aquel en el que todos los elementos de la población tienen una probabilidad conocida de formar parte de la muestra. En el segundo se desconoce o es indeterminable esa probabilidad, restándole ciertos efectos de representatividad.

El muestreo probabilístico ofrece seguridad y control en la forma de la muestra, dando mayor garantía a las pruebas estadísticas que se realizan para explicar el comportamiento de las variables o sus niveles de relación entre sí.

Para realizar la extracción de una muestra probabilística, es indispensable conocer el tamaño de la población, y que cada uno de los elementos que la conforman se encuentre identificado y localizable, así como poseer una herramienta que permita realizar la selección con la ausencia de emociones, sentimientos, gustos y otros aspectos de carácter personal en el investigador, dejando la absoluta responsabilidad de la selección a la metodología empleada.

A continuación se presenta la clasificación del muestreo probabilístico:

- Muestreo aleatorio o al azar.
- Muestreo sistemático.
- Muestreo estratificado.
- Muestreo polietápico.

El seguir o utilizar alguno o algunos de ellos en un proceso de investigación, dependerá principalmente de las características poblacionales así como del objetivo y alcances del estudio.

#### **1.2.2.1 Muestreo aleatorio o al azar.**

Se caracteriza porque es aquel en el que todos los elementos de la población tienen la misma probabilidad de ser seleccionados para la muestra. Para la selección se sugiere observar, que la población sea finita y que cada uno de sus elementos estén identificados, habitualmente por un número.

Finalmente se requiere disponer de algún instrumento que facilite la selección, como puede ser una calculadora científica, una computadora o una tabla de dígitos aleatorios, como la que se presenta en el anexo 1.

Realizar una selección aleatoria resulta impráctica, a pesar del uso de la calculadora o la computadora misma, en la medida en que la población sea muy grande.

### **1.2.2.2 Muestreo sistemático.**

Aún formando parte del muestreo probabilístico, la oportunidad de que los elementos de la población tengan la misma probabilidad de ser seleccionados resulta relativa debido a que una vez seleccionado el primer elemento para la muestra, podría decirse que automáticamente ya se conoce quiénes son los elementos que estarán en la muestra.

Este tipo de muestreo consiste en dividir el tamaño de la población entre el tamaño de la muestra, lo que significa que la población es dividida en tantas partes como el tamaño de la muestra sea y de cada una de esas partes se extraerá un elemento que habrá de integrar a la muestra.

Por ejemplo, si se tiene una población de 500 elementos y se determinó seleccionar a 50 de ellos para obtener el tamaño de la muestra, se divide 500 entre 50, dando como resultado  $k=10$ , lo que indica que la población ha sido fragmentada en 50 partes y de cada una de ellas se tomará un elemento para la muestra, con lo que al final se tendrá el total de los elementos requeridos para el análisis muestral.

Una vez identificados los 50 grupos de 10 elementos cada de uno de ellos, según el ejemplo; se extrae aleatoriamente del primero de los grupos formados un elemento para la muestra. Su póngase que los elementos están identificados con la numeración del 1 al 500, por lo tanto el primero de esos grupos incluye a los elementos enumerados del 1 al 50, el segundo incluirá a aquellos cuyo números estén del 51 al 100, el tercero del 101 al 150 y así sucesivamente hasta el grupo 50 que contiene a los elementos identificados del 451 al 500. Para la primera selección se toma alguno de los instrumentos señalados anteriormente y se busca un número entre 1 y 50, rango del primer grupo. Supóngase que el primer número que aparece en la calculadora o en la tabla de dígitos aleatorios es el 25, lo que indica que el primer elemento de la muestra será el 25, se encuentra en el primer grupo, y a partir de él cuéntense 50 elementos y extráigase el segundo, en este caso será el 75; después de éste cuente otros 50 y seleccione al siguiente, será entonces el 125, y así sucesivamente.

Este tipo de muestreo tiene la ventaja de garantizar la forma de la población, ya que recorre todos los sectores de la misma, esto es, observa desde el primero hasta el último.

La desventaja en este tipo de muestreo, está en los casos en que se dan periodicidades en la distribución de la población, ya que al elegir a los miembros de la muestra con una periodicidad constante ( $k$ ) se puede introducir una homogeneidad que no se da en la población. Supóngase que se está seleccionando una muestra sobre listas de 10 individuos en los que los 5 primeros son varones y los 5 últimos mujeres, si se utiliza el muestreo aleatorio sistemático con  $k=10$  siempre serán seleccionados o sólo hombres o sólo mujeres, no podría haber una representación de los dos sexos.

### **1.2.2.3 Muestreo Estratificado.**

Consiste en considerar características heterogéneas, llamadas estratos o subconjuntos, en la población y que al mismo tiempo son distinguidas por la homogeneidad respecto a alguna característica como la profesión, el sexo, el estado civil, la gerencia a la que se pertenece, entre otras.

Lo que se pretende con este tipo de muestreo es asegurarse de que todos los estratos de interés estarán representados adecuadamente en la muestra. Cada estrato funciona independientemente, pudiendo aplicarse dentro de ellos el muestreo aleatorio simple o el sistemático; para elegir los elementos concretos que formarán parte de la muestra. En ocasiones las dificultades que plantean son demasiado grandes, pues exige un conocimiento detallado de la población: sexos, edades, entidad federativa, campus.

Este tipo de muestreo tiene la ventaja de que permite establecer controles cualitativos para garantizar la representatividad, por lo menos en forma, proporcional de la muestra según la distribución poblacional.

#### **1.2.2.4 Muestreo polietápico (varias etapas).**

Cuando una población presenta tanto detalle de heterogeneidad en sus sectores o estratos que la conforman, es recomendable pasar a un segundo plano o hasta el que sea necesario de clasificación o estratificación y al final realizar la selección aleatoria o sistemática, con las recomendaciones anteriores.

Ejemplo: Se desea extraer una muestra cuyo interés radica en observar a los empleados de una compañía considerando la división a la que pertenece, su género, categoría y turno de trabajo, luego entonces se recomienda estratificar primeramente por división y dentro de cada división hacerlo por género; a partir de esta última variable estratificar por categoría y en ella por turno. En este último nivel de estratificación, se hará la selección aleatoria o sistemática, con la intención de garantizar la representatividad proporcional de la muestra, ya que ésta ofrece mejores condiciones para el análisis.

Como puede deducirse, este tipo de muestreo aparenta ser complicado; sin embargo, ofrece grandes ventajas cuando se dispone de un mayor tiempo y sobre todo si la población evidencia distintos niveles de heterogeneidad o cuando se requiere observar con mayor detalle la distribución poblacional.

#### **1.2.3 TIPOS DE MUESTREO NO PROBABILÍSTICO.**

Circunstancialmente, para estudios exploratorios, el muestreo probabilístico resulta excesivamente costoso y el investigador se asiste con recursos no probabilísticos, aun siendo conscientes de que no sirven para realizar generalizaciones, pues no se tiene certeza de que la muestra extraída sea representativa, ya que no todos los sujetos de la población tienen la misma probabilidad de ser elegidos, debido a que generalmente son seleccionados de acuerdo a criterios, en muchas ocasiones poco ortodoxos.

### **I.2.3.1 Muestreo por Cuotas.**

Se realiza generalmente sobre la base del conocimiento, a veces parcial, de los estratos de la población o de los individuos más representativos o para los fines de la investigación. Tiene semejanza con el muestreo aleatorio estratificado, pero sin el carácter de aleatoriedad de aquél.

En este tipo de muestreo se fijan cuotas que consisten en un número de individuos que reúnen determinadas condiciones, por ejemplo: 20 individuos de 15 a 20 años, de sexo femenino que estudien en escuela pública. Una vez determinada la cuota se eligen los primeros en la población que se encuentren y que cumplan esas características. Este método se utiliza mucho en las encuestas de opinión.

### **I.2.3.2 Muestreo intencional.**

Consiste en que el investigador selecciona directa e intencional mente los individuos de la población. El caso más frecuente de este procedimiento es el utilizar como muestra los individuos a los que se tiene fácil acceso, por ejemplo, las personas que se encuentran descansando en un parque o esperando en la fila del supermercado.

### **I.2.3.3 Bola de nieve.**

Consiste en entrevistar a individuos, los cuales conducen a otros, y así sucesivamente hasta conseguir la muestra requerida. Este tipo de muestreo se emplea cuando se hacen estudios con poblaciones difíciles de acceder a ellas como son las de determinados tipos de enfermos.

#### **I.2.3.4 Muestreo Discrecional.**

A criterio del investigador los elementos son seleccionados si cree que pueden aportar datos o información al estudio, recuérdense las entrevistas de los reporteros de televisión.

#### **I.2.4 DETERMINACIÓN DEL TAMAÑO DE LA MUESTRA.**

Hasta este momento, el lector se habrá dado cuenta del detalle que implica diseñar muestras para una investigación. No obstante, ese detalle es superado cuando existe involucramiento en el área de oportunidad en estudio, así como el conocimiento de lo que se pretende realizar y el objetivo a alcanzar. Ciertamente, las etapas abordadas anteriormente ofrecen al investigador las herramientas, sugiriendo la aplicación de diversos criterios, para la selección de los individuos que formarán parte de la muestra; sin embargo, la importancia de esas etapas se minimiza sin la consideración del número de sujetos que habrán de ser observados, de allí que surge una etapa más para el diseño de la muestra: la determinación de su tamaño.

También la experiencia indica la arbitrariedad con la que se fija el número de elementos que formarán parte de la muestra. Se escucha por los pasillos de las áreas de investigación, por las comunidades e incluso por las oficinas de empresas dedicadas a la investigación de mercados, encuestas y estudios de opinión, que deberá tomarse el 10% ó el 20%, por citar algunos, como tamaño de muestra; cifras que pueden estar sustentadas en muchos años o proyectos de experiencia, soslayando que cada proyecto, estudio, población, universo y muestra son diferentes en tiempo, fondo y forma; por lo que deben ser tratados de manera diferente ya que no aplican reglas generales para ello.

Antes de determinar el tamaño de una muestra, se recomienda considerar lo siguiente: si la población es grande, naturalmente la muestra tiende a ser relativamente pequeña; y si la población es pequeña, entonces la tendencia de la muestra es a ser proporcionalmente grande. A partir de este criterio de observación, se presentan los tres elementos básicos para la determinación del tamaño de una muestra en la investigación social, tanto para

poblaciones finitas como para poblaciones infinitas.

Esos tres elementos son el nivel de confianza, el margen de error y la dispersión en la población, considerada por diversos autores como la variabilidad o varianza. Se refiere como dispersión en la población considerando que en la investigación social, principalmente inciden o se presentan fenómenos de carácter cualitativo donde se aprecia la proporción o porcentaje de elementos que poseen las características o se ven inmersos en los aspectos principales de la investigación.

Para el cálculo del tamaño de una muestra deberá considerarse el tipo de población de donde será extraída, y así seguir alguna de las siguientes fórmulas básicas:

$$\text{Población finita: } n = \frac{N p q Z^2}{E^2 (N - 1) + p q Z^2} \quad (\text{Fórmula 1})$$

$$\text{Población infinita: } n = \frac{Z^2 p q}{E^2} \quad (\text{Fórmula 2})$$

Donde: **Z** Representa el nivel de confianza.

**E** Indica el margen de error o nivel de precisión.

**P** Es la proporción o porcentaje de elementos en la población que tienen la característica en estudio.

**q** Indica el complemento de p:  $q = 1 - p$

**N** Tamaño de la población.

#### **I.2.4.1. Nivel de Confianza.**

De manera sencilla, se desea determinar cuántos de cada 100 elementos que se tomen como muestra, deberán poseer las características de la población. En otros términos es la probabilidad de que la estimación efectuada se acerque a la realidad.

Obviando el detalle de acceso a la tabla de áreas bajo la curva normal, en la cual se obtienen los coeficientes que representan el nivel de confianza en el cálculo del tamaño de la muestra y considerando que comúnmente en ciencias sociales los niveles mayormente utilizados son del 90%, 95% Y 99%, éstos son  $Z=1.64$ ,  $Z=1.96$  y  $Z=2.57$  respectivamente. Más adelante se aplicarán en un ejemplo.

#### **I.2.4.2 Margen de error.**

También llamado error muestral o nivel de precisión, es la diferencia máxima que el investigador está dispuesto a aceptar entre un estadístico y su parámetro. En una investigación siempre se considera la omisión de errores en el proceso, pero la naturaleza de la investigación indicará hasta qué grado es posible su presencia y su tolerancia. Un estadístico será más preciso en la medida que su error sea más pequeño o su precisión sea mayor.

#### **I.2.4.3 Dispersión en la población o variabilidad.**

Antes de iniciar una investigación, el equipo de proyecto deberá tener claro qué tipo de población estudiará, qué características sobresalen, cuál es la variable principal u objeto de estudio, y otros aspectos que son determinantes para el involucramiento en el tema y consecuentemente para realizar diagnósticos y análisis con mayor confiabilidad y credibilidad.

Un factor importante para determinar el tamaño de una muestra es conocer cómo se encuentra distribuida la población en relación con la variable principal de la investigación. Este factor puede ser observado en términos relativos o porcentuales y aplicarse de cualquier manera en la fórmula para su cálculo, siempre y cuando se aplique en el mismo sentido: relativo o porcentual: siempre proporciones o siempre porcentajes.

Si la característica principal del estudio la posee el 70% de la población, a este valor le llamaremos variabilidad o dispersión en la población, y al porcentaje de sujetos que no tienen esa característica 30%, la identificaremos como el complemento a la variabilidad, identificados como **p** y **q** respectivamente. Ambos tienen que sumar 100%.

En caso de que se desconozca esa característica, deberá asignarse la máxima variabilidad correspondiendo el 50% a **p** y el 50% restante a **q**.

Con el conocimiento adquirido hasta el momento, ya es posible determinar el tamaño de una muestra, presentando a continuación algunos ejemplos.

**Ejemplo 1.** De qué tamaño será una muestra si se tiene una población de 500 individuos objetos de estudio, requiriéndose para ello un nivel de confianza del 90% con un margen de error muestral del 8% y se sabe que el 70% de la población posee la característica en estudio, que puede ser el alto nivel de desempeño, contra el 30% con un bajo nivel de desempeño.

Los valores a sustituir en la fórmula 1 son:  $Z = 1.64$  para el 90% de confianza,  $E = 8$  como el porcentaje del error muestral, **p = 70** como valor de la variabilidad, **q = 30** como complemento de la variabilidad y  $N = 500$  como tamaño de la muestra.

Al sustituir los datos en la fórmula 1 y realizar las operaciones tendremos lo siguiente:

$$n :: \frac{(500)(70)(30)(1.64)^2}{(8)^2(500-1) + (70)(30)(1.64)^2}$$

$$n = 75$$

**Ejemplo 2.** Para realizar una encuesta telefónica de opinión entre jóvenes mayores de 18 años en la Ciudad de México, se requiere una muestra con el 95% de confianza, un margen de error muestral del 5% y dadas las características de la población, considerada para este caso como infinita, se tomará la máxima variabilidad.

Bajo el mismo principio de asignación de valores que en el ejemplo 1, al sustituir los datos en la fórmula 2 y realizar las operaciones correspondientes, deberá tomarse una muestra de:

$$n :: \frac{(1.96)^2(50)(50)}{(5)^2}$$

( Fórmula 2 )

$$n = 384$$

**Actividades:**

1. Para confirmar los criterios que se indican respecto a la presencia de poblaciones grandes con tendencia de muestras pequeñas y de poblaciones pequeñas con muestras grandes; calcule 5 muestras con el 95% de confianza, 5% de margen de error y variabilidad del 80%, para las siguientes poblaciones:  $N_1 = 15000$ ,  $N_2 = 10000$ ,  $N_3 = 5000$ ,  $N_4 = 1000$  Y  $N_5 = 350$

Compare el porcentaje de elementos que deberán tomarse como muestra, en cada caso, respecto al tamaño de su población.

¿Qué opinión tiene después de haber realizado los ejercicios?

## **Ejercicio de aplicación**

Los datos que se presentan a continuación, corresponden a la evaluación del desempeño anual de un grupo de 320 trabajadores de la industria de la panadería. Ese puntaje se obtuvo a través de aplicar un instrumento para conocer la opinión de jefes inmediatos, subordinados, compañeros en línea y autoevaluación de cada uno de los trabajadores.

Para determinar el programa de estímulos y compensaciones para el próximo ejercicio, la empresa y el sindicato acuerdan en analizar inicialmente el puntaje obtenido por los trabajadores durante el último año y posteriormente tomar las medidas que beneficien las relaciones contractuales y plan de beneficios para los trabajadores.

Por razones de tiempo, se decide diseñar una muestra de la tabla de resultados de la evaluación del desempeño de los trabajadores y presentar el informe correspondiente a ambas instancias.

*Para ello, se le solicita a usted como consultor en investigación y desarrollo, determinar el tamaño de la muestra que será analizada y seleccionarla sistemáticamente.*

Los datos de los trabajadores son los siguientes y los identificaremos como sigue:

Cada línea contiene 10 datos, cada uno de ellos es el puntaje obtenido por un trabajador, significa que tenemos 10 trabajadores por renglón de la tabla. La primera línea contiene a los trabajadores cuyo número de identificación es, por favor vea la tabla: 1 el primero ( 60) y 10 el último ( 98 ). La segunda línea contiene a los trabajadores identificados con los números del 11 ( 90 ) al 20 ( 74 ) y así sucesivamente hasta el trabajador con el número de identificación 320, cuyo puntaje es 100.



Se requiere que la muestra ofrezca el 95% de confianza (  $Z = 1.96$  ) con un margen de error del 7% (  $E = 7$  ). Observaciones de años anteriores indican que el 80% de los trabajadores son categorizados como de alto desempeño, por lo tanto tomaremos la variabilidad del 80% (  $P = 80$  ) Y su complemento de 20% (  $q = 20$  ).

Sustituyendo en la fórmula 1 para poblaciones finitas tenemos:

$$n = \frac{(320)(80)(20)(1.96)^2}{(7)^2(320 - 1) + (80)(20)(1.96)^2} = 90$$

Lo que indica que se deberá tomar una muestra de 90 trabajadores y seleccionarlos sistemáticamente.

Para la selección sistemática, dividiremos el total de la población  $N = 320$  entre el tamaño de la muestra  $n = 90$

$$K = 320 / 90 = 3.55 \quad (\text{redondearemos a } 4)$$

Esto significa que el primer elemento de la muestra será extraído aleatoriamente de los primeros cuatro de la población. Para seleccionar el segundo, contaremos cuatro elementos y el cuarto será el que debe ser extraído y así sucesivamente de cuatro en cuatro hasta recorrer toda la población o conjunto de trabajadores, en este ejemplo.

Al utilizar la tabla de dígitos aleatorios del anexo 1, observamos que el primer dígito que aparece entre 1 y cuatro es precisamente el 4. Observe la tabla de los datos del ejercicio y el cuarto trabajador tiene una evaluación de 95 puntos. A partir de éste, cada cuatro seleccionará al siguiente, hasta el final del conjunto.

En la siguiente tabla se han marcado los elementos que forman parte de la muestra y deberán ser analizados para el informe que se presentará a la empresa y al sindicato conjuntamente.

60	47	82	95	88	72	67	66	68	98
90	77	86	58	64	95	74	72	88	74
77	39	90	63	98	88	49	42	98	70
39	76	84	91	51	75	50	69	56	100
60	47	82	95	88	72	67	66	68	98
90	77	86	58	64	95	74	72	88	74
77	39	90	63	98	88	49	42	98	70
93	76	84	91	51	75	50	69	56	100
60	47	82	95	88	72	67	66	68	98
90	77	86	58	64	95	74	72	88	74
77	39	90	63	98	88	49	42	98	70
39	76	84	91	51	75	50	69	56	100
60	47	82	95	88	72	67	66	68	98
90	77	86	58	64	95	74	72	88	74
77	39	90	63	98	88	49	42	98	70
39	76	84	91	51	75	50	69	56	100
60	47	82	95	88	72	67	66	68	98
90	77	86	58	64	95	74	72	88	74
77	39	90	63	98	88	49	42	98	70
39	76	84	91	51	75	50	69	56	100
60	47	82	95	88	72	67	66	68	98
90	77	86	58	64	95	74	72	88	74
77	39	90	63	98	88	49	42	98	70
39	76	84	91	51	75	50	69	56	100
60	47	82	95	88	72	67	66	68	98
90	77	86	58	64	95	74	72	88	74
77	39	90	63	98	88	49	42	98	70
39	76	84	91	51	75	50	69	56	100
60	47	82	95	88	72	67	66	68	98
90	77	86	58	64	95	74	72	88	74
77	39	90	63	98	88	49	42	98	70
39	76	84	91	51	75	50	69	56	100

Los datos marcados con color, son los correspondientes a la muestra seleccionada sistemáticamente. Debido a que se redondeó de 3.55 a 4, el total de datos de la muestra fue finalmente 63. La diferencia es llamada error de redondeo por muestreo sistemático.

### **Actividad adicional**

Con los datos de la muestra, elabore una tabla de distribución de frecuencias en 6 intervalos de clase. Dibuje su histograma y calcule la media, desviación estándar, sesgo y curtosis. Interprete los resultados y presente el informe requerido. Si es necesario, calcule algunos percentiles que apoyen su exposición.

En resumen, esta unidad le ha proporcionado los elementos necesarios para el diseño de muestras, recordando que cada etapa es igualmente importante. Al calcular el tamaño de la muestra, deberá procederse a su selección de acuerdo al método y tipo determinado por el investigador. Posteriormente habrá que realizarse el procedimiento para la organización de los datos, realización de pruebas estadísticas y todo aquello relacionado con los objetivos de la investigación.

## **UNIDAD 2**

### **HIPÓTESIS ESTADÍSTICAS**

#### **INTRODUCCIÓN**

La investigación social parte de suposiciones adoptadas para explicar hechos a través de la relación entre variables. Para su tratamiento, esos supuestos son traducidos a términos estadísticos y sometidos a pruebas de validez, por lo que el investigador deberá identificar claramente el tipo de variables en estudio, así como la relación que guardan entre sí.

Esta unidad presenta la metodología para el diseño y contrastación de hipótesis, así como los errores que pueden cometerse al tomar una decisión estadística y las condiciones en que éstas deberán tomarse.

#### **OBJETIVO**

Conocer la estructura y diseño de hipótesis estadísticas, su aplicación en la investigación social y su importancia en la toma de decisiones.

#### **TEMARIO**

##### **II. HIPÓTESIS ESTADÍSTICAS.**

###### **II.1. Hipótesis Estadísticas.**

###### **II.1.1 Características de las hipótesis.**

###### **II.2. Hipótesis de Trabajo.**

###### **II.3. Contraste de Hipótesis.**

###### **II.4. Hipótesis Nula.**

###### **II.5. Hipótesis Alternativa.**

II.6. Errores Estadísticos.

II.7. Método para la comprobación de hipótesis estadísticas.

## **II.1. HIPÓTESIS ESTADÍSTICA.**

Enunciado que afirma la relación, comportamiento, tendencia y diversos aspectos de ocurrencia entre dos o más variables de una muestra o población y está sujeto a comprobación. Este enunciado podrá ser verdadero o no verdadero. En otros términos, podrá aceptarse o no su validez.

### **II.1.1 Características de las hipótesis**

- Deben redactarse claramente, de tal manera que no permita interpretaciones.
- Deben ser objetivas sin incluir juicios de valor.
- Las variables deben estar implícitas o explícitas en la redacción.
- Deben estar relacionadas con técnicas estadísticas para someterlas a prueba.
- Deben ser estadísticamente comprobables.

Ejemplo: La causa principal del desempleo en México se debe a la mala suerte de los mexicanos.

Esta afirmación ciertamente es un supuesto que puede ser verdadero o no; sin embargo, no es posible medir el concepto mala suerte, por lo que esta hipótesis carece de comprobación científica.

## **II.2. HIPÓTESIS DE TRABAJO.**

Es aquella que plantea el investigador o la investigación como objeto o base de su análisis y estudio; trata de dar una explicación tentativa al fenómeno que se está investigando. Esta es la hipótesis que el investigador tratará de aceptar como resultado de su investigación, rechazando la hipótesis nula, la cual se comentará en otro apartado.

Ejemplo: La causa principal del desempleo en México se debe principalmente a la reducción de las exportaciones, lo que ha provocado el cierre de fuentes de trabajo.

En este segundo caso, sí es posible medir la relación entre la reducción de las exportaciones y cierre de fuentes de trabajo con las tasas del desempleo y con fundamento en los resultados llegar a la conclusión de validar o no la hipótesis mencionada.

## **II.3. CONTRASTE DE HIPÓTESIS.**

Es un proceso estadístico que permite elegir una hipótesis de trabajo entre dos posibles y contraria una a la otra. El contraste inicia con la formulación de dos hipótesis sobre el valor de algún parámetro o característica poblacional, siendo ambas contradictorias, si una de ellas es verdadera por consecuencia la otra no lo será.

En el caso del segundo ejemplo, los planteamientos serían:

- La tasa del desempleo está determinada por la reducción de las exportaciones
- La tasa del desempleo no está determinada por la reducción de las exportaciones

Al analizar los datos de la muestra seleccionada para la comprobación de la hipótesis, se concluirá cuál de las dos es verdadera. Si la primera no lo es, la conclusión será que existen otras causas o variables, diferentes a la reducción de exportaciones, que estén provocando el fenómeno del desempleo.

El contraste de hipótesis puede ser bilateral o unilateral, será bilateral cuando la hipótesis esté planteada o diseñada en términos de relación entre variables, recordando que la relación entre dos o más variables puede ser de igualdad ( = ), una mayor o menor que la otra ( < , > ).

Si dos valores no son iguales, al ser diferentes podrá darse el caso de que sea mayor o menor uno que el otro, surgiendo aquí el concepto de hipótesis unilateral o bilateral.

***Ejemplo de hipótesis unilateral:***

Al recibir el título profesional, los empleados obtienen mayor salario que aquellos que no han concluido sus estudios universitarios.

En este caso la hipótesis sólo podrá ser rechazada si el salario de los titulados no es mayor que los que no poseen el nivel licenciatura.

***Ejemplo de hipótesis bilateral:***

Los puestos vacantes son ocupados por personas de 20 a 25 años de edad.

Al observar la selección del personal, la hipótesis sólo será rechazada si algún aspirante es seleccionado con menos de 20 años o mayor a los 25 años de edad.

**II.4. HIPÓTESIS NULA (  $H_0$  ).**

Para realizar el contraste de hipótesis se supone cierta una de ellas, llamada **hipótesis nula**  $H_0$ , la cual es formulada por el investigador con el propósito de ser rechazada, este enfoque resulta en ocasiones incomprensible, ya que se cuestiona ¿cómo es posible que se plantee algo que pretende rechazarse?

Con un ejemplo se explicará. La industria electrónica japonesa somete sus aparatos a estrictas pruebas para el control de calidad. De un lote de DVD's seleccionan una muestra aleatoria y las introducen en cámaras de prueba para someterlas al tratamiento promedio que les da el usuario, como usted y yo; esto es, se exponen a caídas, temperatura, calor y derrame de líquidos. Al finalizar la prueba se encienden todos los DVD's que fueron sometidos a control y si falla por lo menos uno de ellos, todo el lote es destruido, no sólo los que participaron en la prueba; de allí la alta calidad de los productos japoneses. ¿Entendemos ahora a los investigadores?

Para entenderlos mejor, le diré que la hipótesis nula se plantea contraria a la hipótesis de trabajo y se formula en términos estadísticos.

**Ejemplo de hipótesis de trabajo:**

Asisten a los programas de capacitación hasta el 30% de los trabajadores de la empresa.

Hipótesis nula:  $P > 30\%$

Obsérvese la oposición que presenta la hipótesis nula en relación con el planteamiento de la hipótesis de trabajo y el término proporcional ( P ) que se emplea para su formulación.

**II.5. HIPÓTESIS ALTERNATIVA ( H1 ).**

Por su parte, la hipótesis alternativa se opone a la hipótesis nula y está en relación directa con el planteamiento de la hipótesis de trabajo, esto es, tienen el mismo sentido aunque planteada la primera en términos estadísticos y la última en el lenguaje de la investigación.

Retomando el ejemplo anterior, se formularán las hipótesis mencionadas:

*Hipótesis de trabajo:* "Asisten a los programas de capacitación hasta el 30% de los trabajadores de la empresa"

Hipótesis Nula:  $H_0: P > 30\%$   
Hipótesis alternativa:  $H_1: P < = 30\%$  El símbolo (  $< =$  ) léase menor o igual

Nótese que la formulación de la hipótesis alternativa indica lo mismo que la hipótesis de trabajo y ambas se oponen a la hipótesis nula.

La prueba de hipótesis se realiza con el planteamiento de la hipótesis nula (  $H_0$  ), lo que significa que la que se acepta o no se acepta es ésta. Si  $H_0$  no es aceptada, el investigador asume que no se tiene evidencia suficiente para aceptarla, confirmando por consecuencia su hipótesis de trabajo.

## **II.6. ERRORES ESTADÍSTICOS.**

Al realizar pruebas para la contrastación de hipótesis, finalmente deberá tomarse la decisión de aceptar o no aceptar la hipótesis nula (  $H_0$  ), que no garantiza que la conclusión a la que se ha llegado sea realmente lo que esta ocurriendo en el contexto poblacional. El investigador puede errar el planteamiento de sus hipótesis de trabajo o el diseño de la hipótesis estadística, razón por la que deberán considerarse la posible comisión de errores al tomar la decisión o los niveles de incertidumbre a los que están sometidas las pruebas de hipótesis. Estos errores son clasificados e identificados en estadística inferencial como el Error Tipo I y el Error Tipo II

### **Error Tipo I**

El Error Tipo I se comete cuando en la decisión se rechaza la Hipótesis Nula (  $H_0$  ) siendo ésta verdadera.

### **Error Tipo II**

El Error Tipo II se comete cuando en la decisión se acepta la Hipótesis Nula (  $H_0$  ) siendo ésta no verdadera.

Pero realmente ¿qué importancia puede tener el saber si se tomó una decisión equivocada o no? La importancia se sostiene en la trascendencia que puede tener para la investigación o proyecto el saber que la explicación de los fenómenos hecha por el estudio, no corresponde a las causas reales de la relación entre las variables analizadas, lo que consecuentemente provocaría decisiones organizacionales irreales.

Sin embargo, el saber que las decisiones estadísticas están soportadas por altos niveles de certidumbre en el análisis de la relación causa-efecto entre sus variables, obviamente que conducirá a tomar mejores y más oportunas decisiones.

## **II.7. MÉTODO PARA LA COMPROBACIÓN DE HIPÓTESIS.**

Para la comprobación de hipótesis estadísticas se sigue un método de 7 pasos, sencillo y práctico, que facilita el análisis de relación entre variables, pruebas de independencia y otros aspectos de interés para la investigación y la toma de decisiones.

### **Paso 1. Planteamiento de la hipótesis de trabajo**

En ella el investigador trata de dar explicación a la relación entre las variables en estudio. Se plantea en formato de texto.

**Ejemplo:** "El desempeño de los trabajadores en el campo laboral, es una de las consecuencias de su rendimiento en los programas de capacitación".

### **Paso 2. Planteamiento de la hipótesis estadística.**

Se identifican las variables y su relación en la hipótesis de trabajo y se traduce a términos estadísticos, creando con ello las hipótesis nula (  $H_0$  ) e hipótesis alternativa (  $H_1$  ).

**Ejemplo:**  $H_0 : Q = 0$

$$H1 : Q \neq 0$$

**Paso 3. Determinación del estadístico de prueba**

Se especifica la fórmula aritmética que será utilizada para realizar el cálculo que sustentará la prueba.

**Ejemplo:**

$$Q = \frac{D - BC}{AD + BC}$$

Este estadístico es llamado Q de Kendall y permite correlacionar dos variables nominales de dos categorías cada una. Considérese que en este ejemplo, las categorías de la variable "Desempeño" son "Buen desempeño" y "Mal desempeño"; y para la variable "Rendimiento en los programas de capacitación" son "Buen rendimiento" y "Mal rendimiento".

**Paso 4. Planteamiento de la regla de decisión**

Se especifican las condiciones en las que se deberá aceptar o no aceptar la hipótesis nula.

**Ejemplo:**

1. Aceptar  $H_0$  si el valor de  $Q < 0.8$
2. No aceptar  $H_0$  si el valor de  $Q \geq 0.8$

**Paso 5. Cálculo del estadístico de prueba**

Los datos de la muestra son sustituidos en la fórmula aritmética para obtener el resultado del cálculo.

**Ejemplo:**

Supóngase que al sustituir los valores y calcular el estadístico de prueba se obtiene el valor de  $Q = 0.79$

**Paso 6. Decisión estadística.**

El resultado obtenido del cálculo anterior se compara en la regla de decisión y se indica si se acepta o no se acepta la hipótesis nula.

**Ejemplo:** Comparando el valor calculado de Q en el paso 4, se toma la decisión de aceptar  $H_0$ .

**Paso 7. Conclusiones de la prueba.**

En los términos en que fue planteada la hipótesis de trabajo se indican los resultados de la prueba y la evidencia que presentó para ser aceptada o no aceptada.

**Ejemplo:** Con los datos de la muestra se observa evidencia de que el desempeño de los trabajadores no es un factor determinado por el rendimiento en los programas de capacitación.

## **UNIDAD 3**

### **PRUEBAS ESTADÍSTICAS DE CORRELACIÓN E INDEPENDENCIA ENTRE VARIABLES.**

#### **INTRODUCCIÓN**

Por medio de las pruebas estadísticas para la comprobación de hipótesis se pretende encontrar la relación causal entre las variables, así como su fuerza, magnitud y sentido.

Las pruebas estadísticas ofrecen evidencias para determinar la fuerza de la relación. Para ello se considera el tipo de variables por nivel de medición en virtud de que existen pruebas específicas para cada uno de ellos.

En esta unidad se presentan pruebas de correlación para variables numéricas y no numéricas, así como para variables nominales y ordinales, con las que el investigador adquirirá los elementos y criterios básicos para la comprobación de sus hipótesis de trabajo.

#### **OBJETIVO**

Conocer e identificar las pruebas estadísticas para la contrastación de hipótesis en la investigación social.

#### **TEMARIO**

#### **III PRUEBAS ESTADÍSTICAS DE CORRELACIÓN E INDEPENDENCIA ENTRE VARIABLES.**

III.1. Conceptos.

III .2. Prueba de correlación Q de Kendall.

III.3. Prueba de correlación R de Pearson.

III.4. Prueba de independencia entre variables Ji Cuadrada.

III.5. Prueba de correlación C de Pearson.

Fundamentalmente, toda investigación busca explicar el comportamiento de los fenómenos en función de la relación entre variables. Teórica o empíricamente, el investigador supone cuáles son las posibles causas que originan que una variable determine la fuerza de relación entre dos o más de ellas; sin embargo, esas suposiciones deberán sustentarse mediante la aplicación de diversas pruebas estadísticas que le permitan confirmar o desechar las hipótesis que se ha planteado para su proceso de investigación.

En este apartado se incluyen las pruebas estadísticas básicas para la comprobación de hipótesis de relación entre variables, las que permitirán argumentar, a partir de los datos observados, cuál es la magnitud de la fuerza de asociación que presentan las variables en estudio. Estas pruebas indicarán, según sea el caso, si dos variables mantienen la relación de dependencia y en qué magnitud, para que el investigador tenga elementos para llegar a conclusiones que le permitirán orientar la toma de decisiones.

### III.1 CONCEPTOS

Es menester que el lector se involucre con algunos términos que se presentarán a lo largo de las siguientes líneas, sugiriéndole asistirse por otras fuentes para finalmente formarse el mejor concepto de los elementos empleados en el proceso para la comprobación o contrastación de hipótesis estadísticas.

**Nivel de significación:** Es la probabilidad de cometer el error tipo 1, esto es, rechazar la hipótesis nula (  $H_0$  ) cuando ésta es verdadera.

**Grados de libertad:**

Número de posibilidades que en un conjunto de datos, se tiene para la asignación de valores, también identificado como la cantidad de variables independientes para la realización de una prueba estadística. Si los grados de libertad de un modelo tienden a incrementarse, la forma de la distribución tiende a ser normal.

**Tabla de congruencia:**

También llamada de doble entrada, es aquella que presenta matricialmente la relación entre dos variables de relación entre dos variables de Esa relación se observa mediante las frecuencias de incidencia entre esos elementos de. Las variables relacionadas. La intersección entre un renglón y una columna es llamada celda.

**Frecuencias marginales**

Es la suma total de las frecuencias absolutas que corresponde a cada valor, intervalo o categoría de una variable que se está relacionando con otra en una tabla de contingencia o de doble entrada.

**Frecuencias conjuntas**

Es la frecuencia absoluta que se observa en cada celda de una tabla de contingencia.

**Frecuencias marginales**

Es la suma total de las frecuencias absolutas que corresponde a cada valor, intervalo o categoría de una variable que se está relacionando con otra en una tabla de contingencia o de doble entrada.

**Frecuencias conjuntas**

Es la frecuencia absoluta que se observa en cada celda de una tabla de contingencia.

**Frecuencias  
observadas:**

Para el análisis de relación entre dos variables, la frecuencia observada es el número de casos que, a partir de los datos en estudio, forman parte de la intersección entre un renglón y una columna en una tabla de contingencia.

**Frecuencias  
esperadas:**

Para el análisis de relación entre dos variables, la frecuencia esperada es el número de casos que de acuerdo a la distribución marginal, la tabla de contingencia esperaba en cada celda para evidenciar la independencia entre variables.

## **111.2 PRUEBA DE CORRELACIÓN Q DE KENDALL**

En ciencias sociales es frecuente que el investigador diseñe sus instrumentos para la recolección de datos, considerando sólo dos características de las variables como: "SI" y "NO", "De Acuerdo" y "En desacuerdo", "Hombre" y "Mujer". Este tipo de respuestas es el factor más elemental de estudio; sin embargo, no por elemental carece de importancia; ya que existen diversas pruebas estadísticas que permiten encontrar las evidencias que justifican el comportamiento de dos variables. Recuerde que las variables en ciencias sociales son medidas en diversas escalas o niveles, según su naturaleza.

La prueba Q de Kendall, permite medir la correlación entre dos variables nominales con dos categorías cada una de ellas, como el sexo: "Masculino" y "Femenino", o como las respuestas a la pregunta de ¿Se encuentra satisfecho con los programas de capacitación de su empresa? Cuyas respuestas son: "Si" y "NO".

Es una prueba sencilla, muchos investigadores y especialistas la consideran como débil y sin importancia; claro que existen pruebas estadísticas de mayor complejidad para variables nominales; sin embargo, es un indicador que permite, con un nivel alto de confianza, observar el comportamiento entre ese tipo de variables, ofreciendo elementos sobre la fuerza en que están interactuando.

Con esta prueba se obtiene el coeficiente Q, el cual parte de la observación de las frecuencias observadas conjuntas entre dos variables nominales de dos categorías. El coeficiente Q tiene un rango de variación de cero a uno, en valor absoluto. Los valores extremos que se pueden obtener al realizar el cálculo correspondiente indican lo siguiente: si el coeficiente es igual a cero, significa que entre las dos variables en estudio, no existe relación alguna. Si el coeficiente es igual a uno, entonces se concluye que las variables se encuentran totalmente relacionadas entre sí, lo que significa que una depende totalmente de la otra. No obstante, en la práctica y sobre todo en ciencias sociales, esos valores extremos son materialmente imposibles de presentarse.

**Ejemplo:** En una organización se pretende conocer el impacto que ha tenido el rendimiento de los empleados en los programas de capacitación con su desempeño en el trabajo. La variable rendimiento está clasificada en "Rendimiento eficaz" y "Rendimiento ineficaz". El desempeño el trabajo está medido como "Buen desempeño" y "No buen desempeño". Supóngase al realizar los cálculos del coeficiente Q de Kendall tenemos un valor de  $Q = 0.7$  lo que indica que las variables nos están totalmente relacionadas, pero si existe cierta dependencia entre ellas. La conclusión en este caso es que el grado de asociación alcanza el 70% de los casos observados, lo que en otras palabras refiere que el desempeño en el trabajo mantiene relación del 70% con el rendimiento en los programas de capacitación.

El coeficiente Q de Kendall se calcula con el siguiente estadístico:

$$Q = \frac{AD - BC}{AD + BC} \quad (3)$$

Donde los coeficientes A, B, C y D son las frecuencias conjuntas observadas en la tabla de contingencia o matriz de doble entrada:

		Variable 1	
		Categoría 1	Categoría 2
Variable 2	Categoría 1	A	B
	Categoría 2	C	D

Ejemplo: con los datos de la siguiente tabla, se requiere probar la hipótesis nula de que la afectividad de los trabajadores no influye en la relación con sus compañeros de trabajo.

		¿Se considera usted afectivo?	
		Sí	No
¿Cómo considera la relación Con sus compañeros de trabajo?	Buena	23	15
	Mala	18	25

**Hipótesis estadística:**

$$H_0 : Q = 0$$

$$H_1 : Q \neq 0$$

Interprete a  $Q = 0$  como la inexistencia de relación entre las variables analizadas, lo que significa que se espera que la afectividad de los trabajadores es independiente de las relaciones en el trabajo.

La regla de decisión será:

1. Aceptar  $H_0$ , si el coeficiente  $Q$  a calcular es menor a 0.8
2. No aceptar  $H_0$  si el coeficiente  $Q$  es mayor o igual a 0.8

Esto significa que el criterio empleado para determinar si la afectividad influye en las relaciones en el trabajo, es que por lo menos en el 80% de los datos observados se confirme el supuesto del investigador, esto es, que no sea aceptada la hipótesis nula. Recuerde que la intención de plantear la hipótesis nula es rechazarla y su sentido es contrario a la hipótesis de trabajo.

Al calcular el coeficiente  $Q$  con los datos de la tabla encontramos que su valor es  $Q = 0.36$  que al compararlo en la regla de decisión nos lleva a la tomar la decisión estadística correspondiente:

Decisión estadística: Aceptar  $H_0$ .

Conclusión: Los datos utilizados para la prueba pretenden confirmar que la afectividad de los trabajadores es independiente, o tiene poca relación o influencia en las relaciones de éstos con sus compañeros; ya que se observó, que sólo 36 de cada 100, si presentan relación alguna ( 36% ) entre ambas variables. Seguramente que las relaciones personales están influidas, por lo menos a partir de los datos de la muestra tomada para la investigación, por otras variables como el liderazgo, la motivación, entre otras.

### III.3 COEFICIENTE DE CORRELACIÓN DE PEARSON.

Como ha sido referido anteriormente, las variables de mayor presencia o incidencia en las ciencias sociales son de carácter no numérico, característica que hace la diferencia con las demás disciplinas o áreas del conocimiento. Actualmente, nos encontramos en la etapa de adecuación a los nuevos ritmos y exigencia de los mercados tanto nacionales como internacionales, sea cual fuere su ámbito o contexto de acción: social, económico, cultural, etc.; para cambiar el paradigma de que sólo en la investigación social se trabaja con datos no numéricos, de allí la discrepancia, también actual, entre investigadores y profesionales de la estadística aplicada a las ciencias sociales.

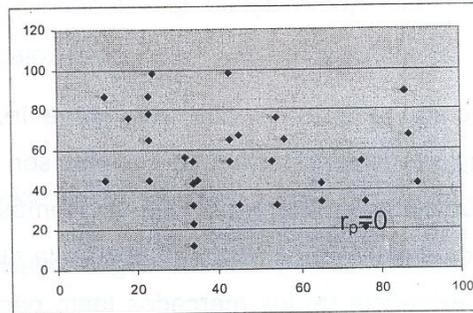
La tendencia actual es que los fenómenos cualitativos sean medidos escalaramente desde el diseño de la investigación, con la ventaja de tener mayor libertad o movilidad para la medición de las variables y su relación entre ellas. Uno de los recursos de mayor confiabilidad para medir qué tanto impacta el comportamiento de una variable con aquellas con quienes interactúa, es el coeficiente de correlación de Pearson, mismo que también mide el grado de asociación entre variables, pero en este caso sólo escalares, categoría en que clasificamos a las variables de nivel intervalar y racional.

Este coeficiente de correlación Pearson se mueve también alrededor del intervalo de 0 a 1. Si el coeficiente toma el valor de 1 se dice que la correlación es perfecta, pero si toma el valor cero se identifica como correlación nula.

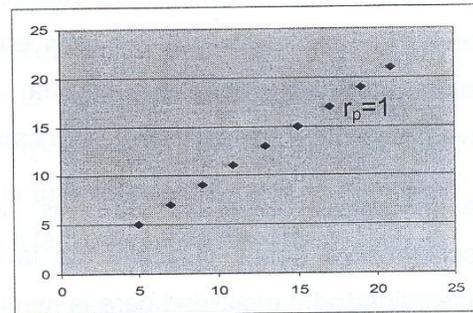
Cabe mencionar que este coeficiente es utilizado sólo con datos numéricos no agrupados por intervalos, o lo que llamamos lista de datos.

La correlación entre variables se determina mediante el coeficiente  $r_p$  de Pearson cuyo rango de variación fluctúa entre 0 y 1 y es aplicable a variables de nivel de intervalo.

El coeficiente de correlación será igual a cero si la correlación es nula. Las variables son independientes.



El coeficiente de correlación será igual a uno si la correlación es perfecta. Las variables son dependientes.



El estadístico de prueba para el coeficiente de correlación de Pearson es el siguiente:

$$r_p = \frac{N \sum XY - (\sum X)(\sum Y)}{\sqrt{[N \sum X^2 - (\sum X)^2][N \sum Y^2 - (\sum Y)^2]}} \quad (4)$$

En esta prueba, la hipótesis nula apunta también, como el caso del coeficiente Q de Kendall, a confirmar la inexistencia de correlación entre las variables numéricas consideradas para el análisis.

Ejemplo: Los siguientes datos corresponden a las horas semanales que un grupo de 10 mujeres que trabajan desde su casa, dedican a su negocio personal y los ingresos promedio obtenidos

Madre	1	2	3	4	5	6	7	8	9	10	
Horas	12	10	8	6	9	12	15	11	20	6	
Ingreso	2400	1700	1900	1200	2200	1800	3000	2400	3800	1000	

Se requiere probar la hipótesis de trabajo de que los ingresos de las personas que trabajan desde su casa, están determinados por las horas que le dedican al negocio personal.

En este ejemplo se desea conocer el nivel de correlación entre dos variables numéricas escalares con datos no agrupados, por lo tanto la hipótesis estadística será la siguiente:

Hipótesis nula  $H_0: r_p = 0$

Hipótesis alternativa  $H_1: r_p \neq 0$

El estadístico de prueba a utilizar es el coeficiente de correlación de Pearson ( 4 ), estructurando la regla de decisión como sigue:

1. Aceptar  $H_0$  si el valor de  $r_p < 0.8 >$
2. No aceptar  $H_0$  si el valore de  $r_p \geq 0.8$

Al sustituir los datos requeridos por el estadístico de prueba y realizar las operaciones respectivas, se obtiene que el coeficiente de correlación es:

$$r_p = 0.93$$

Al comparar este resultado en la regla de decisión, la hipótesis nula no se acepta, lo que indica, por la magnitud del valor obtenido, que existe una correlación muy alta entre las horas de trabajo semanales y el ingreso promedio de las mujeres que laboran desde su hogar.

### III.4 PRUEBA JI CUADRADA ( $\chi^2$ ) DE INDEPENDENCIA ENTRE VARIABLES.

Entre otras aplicaciones, la prueba Ji cuadrada ( $\chi^2$ ) pretende evidenciar, para la relación entre variables nominales u ordinales, qué tan dependientes o independientes son una de la otra, o qué tanto una explica el comportamiento de la otra, por medio de la comparación de la distribución de las frecuencias empíricamente obtenidas en el proceso de recolección de datos, con la distribución de las frecuencias que teóricamente se esperaba por categoría de incidencia.

Un ejemplo de una distribución de frecuencias observadas se presenta en una tabla de contingencia como la siguiente ya partir de ella se realiza la prueba.

Estado civil	Rendimiento en el trabajo			Total
	Bajo	Medio	Alto	
Casado	82	76	20	178
Soltero	75	27	53	155
Total	157	103	73	333

El estadístico utilizado para esta prueba es:

$$\chi^2 = \sum \frac{(\mathbf{Fo} - \mathbf{Fe})^2}{\mathbf{Fe}} \quad (5)$$

Donde:

Fo = frecuencia observada en la incidencia entre dos categorías de las variables.

Fe = frecuencia esperada por la tabla de contingencia entre esas dos categorías.

La prueba consiste en comparar el valor calculado de  $X^2$  con el obtenido a partir de la tabla de la Distribución  $X^2$  que se presenta en el anexo 2. Para acceder a esa tabla deberá determinarse el nivel de significación o el nivel de confianza con el que se desea hacer la prueba y los grados de libertad de la distribución, obtenidos de la siguiente manera:

$$\text{Grados de libertad} = ( \text{total de renglones} - 1 ) ( \text{total de columnas} - 1 ) \quad (6) .$$

La tabla de contingencia que se presenta en este apartado tiene 2 renglones y 3 columnas, por lo tanto tiene  $(2 - 1) (3 - 1) = (1) (2) = 2$  grados de libertad.

Para la obtención del valor de  $X^2$  calculado, la tabla de contingencia presenta las frecuencias observadas. Las frecuencias esperadas se calculan con la siguiente expresión:

$$Fe = \frac{ ( \text{Total de frecuencias del renglón} ) ( \text{Total de frecuencias de la columna} ) }{ ( \text{Suma total de frecuencias marginales de la tabla} ) } \quad (7)$$

En la tabla del ejemplo, la frecuencia observada en la primera celda es 82, que son el total de personas casadas con bajo rendimiento en el trabajo. Si se desea obtener la frecuencia esperada en esa celda, utilicemos la fórmula ( 7 ) Y obtendremos:

$$Fe = \frac{ ( 178 ) ( 157 ) }{ ( 333 ) } = 83.92$$

Significa que para evidenciar que las variables son independientes, la tabla de contingencia esperaba que del total de personas tomadas para la prueba, 83.92 fueran casadas con bajo rendimiento en el trabajo. Nótese que en éste caso la diferencia entre la frecuencia observada (82) y la esperada (83.92) es mínima. Por razones prácticas siga el criterio de redondear sus resultados al entero más cercano, en este caso la frecuencia esperada obtenida se redondea a 4.

En la siguiente tabla se presentan entre paréntesis las frecuencias esperadas o teóricas, para su comparación con las frecuencias observadas o empíricas:

Estado	Rendimiento en el trabajo			Total
	Bajo	Medio	Alto	
Casado	82 <b>(84)</b>	76 <b>(55)</b>	20 <b>(39)</b>	178
Soltero	75 <b>(73)</b>	27 <b>(48)</b>	53 <b>(34)</b>	155
Total	157	103	73	333

La prueba  $X^2$  compara la distribución de esas frecuencias observadas y esperadas por cada celda, y dependiendo qué tan grande se presente la diferencia, saca las conclusiones de si ésta es significativa para determinar si las variables puestas a prueba son independientes o no.

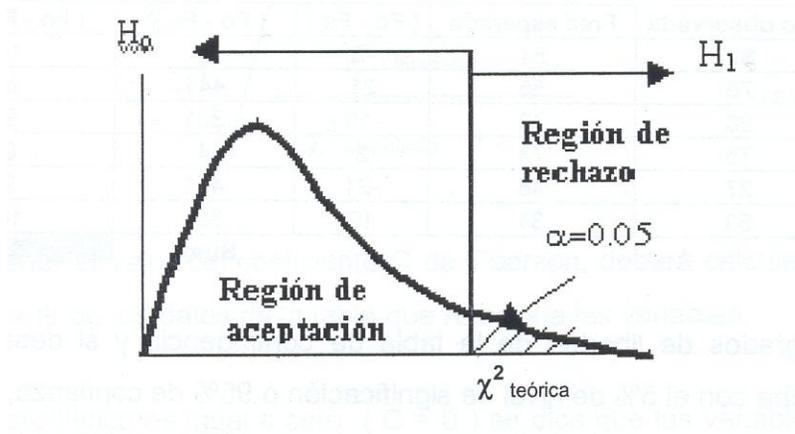
Si la diferencia no es notablemente grande, la conclusión es que las variables relacionadas son independientes y una no explica a la otra, lo que en el ejemplo significaría que el rendimiento en el trabajo no depende del estado civil de los empleados, sino que serán otras causas las que lo determinen. Si por el contrario, las diferencias mencionadas sí son significativamente grandes, la conclusión es que sí existe dependencia entre las variables en estudio, lo que indicaría en el ejemplo que el estado civil determina o influye en el rendimiento laboral.

Analice las diferencias indicadas y le pregunto: ¿Cuál es su conclusión? Continúe con el estudio de los detalles de la prueba y compare su conclusión, suerte.

supuesto que obvia tiempo de trabajo y permite tomar más oportunamente las decisiones; sin embargo, se recomienda que en los primeros ejercicios se actúe con cautela, ya que para determinar si la diferencia entre dos valores es grande o pequeña, dependerá de diversos factores como el contexto y las unidades de medición, entre otros.

Por ejemplo: la diferencia de un minuto entre iniciar una sesión de trabajo antes o después de la hora fijada, tal vez no tenga ninguna trascendencia y se apreciaría como mínima o despreciable. Pero si ese minuto de diferencia se contextualiza en el suministro de oxígeno a un paciente hospitalario, puede ser fatal.

Continuando con la presentación y realización de nuestra prueba  $\chi^2$ , obsérvese la siguiente gráfica, en ella se indica la región de aceptación de la hipótesis nula  $H_0$  la región de no aceptación, también referida como nivel de significación  $\alpha$  ( alfa ), en este caso de 0.05 ó 5%.



La regla de decisión de esta prueba se plantea a partir de indicar cuál es el valor de  $\chi^2$  que determinará la frontera entre la región de aceptación y la región de rechazo. Este valor es llamado  $\chi^2$  teórica y se obtiene de la tabla del anexo 2, considerando los grados de libertad y el nivel de significación con que se realizará la contrastación de la hipótesis nula. El valor de la  $\chi^2$  que se calcula con los datos de la tabla de contingencia, se compara con el de la  $\chi^2$  teórica. Si el valor calculado es menor al valor teórico, entonces

se acepta la hipótesis nula (  $H_0$  ), lo que indica que las variables son independientes; pero si el valor calculado es mayor al valor teórico, entonces está cayendo en la zona de no aceptación, por lo que la hipótesis nula (  $H_0$  ), no es aceptada, refiriendo entonces que las variables no son independientes, esto es que una explica el comportamiento de la otra.

Regla de decisión de la prueba  $X^2$

1. Aceptar  $H_0$  si:  $X^2$  calculada <  $X^2$  teórica

2. No aceptar  $H_0$  si:  $X^2$  calculada  $\geq$   $X^2$  teórica

Para concluir la prueba en nuestro ejemplo, calculemos el valor de \*\*\* los datos de la tabla de contingencia, siguiendo la fórmula ( 5 ):

Celda	Frec observada	Frec esperada	(Fo - Fe)	( Fo - Fe ) <sup>2</sup>	( Fo - Fe) <sup>2</sup> / Fe
1	82	84	-2	4	0.05
2	76	55	21	441	8.02
3	20	39	-19	361	9.26
4	75	73	2	4	0.05
5	27	48	-21	441	9.19
6	53	34	19	361	10.62
Suma					<b>37.19</b>

Con los dos grados de libertad de la tabla de contingencia y si deseamos realizar la prueba con el 5% de nivel de significación ó 95% de confianza, como a usted se le facilite, vea la tabla 2 del anexo y en la primera columna localice 2 grados de libertad y busque la intersección con la columna 0.05, allí encontrará que el valor de la \*\*\* teórica = 5.99147

Al comparar ambos valores ( $X^2$  calculada = 37.19 y  $X^2$  teórica = 5.99147 ) en la regla de decisión, se rechaza la hipótesis nula (  $H_0$  ), concluyendo que las variables no son independientes, lo que se interpreta afirmando que el rendimiento en el trabajo sí mantiene relación con el estado civil de los empleados. ¿Confirmó su decisión anterior?

### III.5 PRUEBA DE CORRELACIÓN C DE PEARSON

En ciencias sociales es común que la investigación requiera de realizar pruebas de correlación entre variables nominales de más de dos categorías, al menos una de ellas. En este caso podrá disponer de un estadístico sencillo que se apoya en el principio del coeficiente  $X^2$ .

El coeficiente C de Pearson es una herramienta que permite medir el nivel de correlación entre dos variables nominales con más de dos categorías como estado civil: soltero, casado, divorciado; como por ejemplo en la pregunta siguiente: ¿está usted de acuerdo en trabajar en fines de semana? Cuyas respuestas son: de acuerdo, ni de acuerdo ni en desacuerdo, en desacuerdo.

El estadístico de prueba es el siguiente:

$$C = \sqrt{\frac{X^2}{X^2 \text{ calculada} + N}} \quad (8)$$

Para obtener el valor del coeficiente C de Pearson, deberá calcularse el valor de  $X^2$  a partir de los datos de la tabla que relaciona las variables.

Si este estadístico es igual a cero ( $C = 0$ ) se dice que las variables no están correlacionadas o son independientes. Entre más cercano se encuentre el coeficiente de cero mayor será la tendencia a que las variables comparadas no guarden relación entre sí. En la medida que el coeficiente vaya aumentado su valor, las variables estarán más correlacionadas. En otras palabras, una de ellas explicará el comportamiento de la otra. El máximo valor que puede tomar el coeficiente C estará cercano a la unidad, indicando la existencia de alto nivel de asociación entre las variables.

Dado que es una prueba de correlación, la hipótesis nula se plantea en los siguientes términos

Hipótesis nula            Ho:      $C = 0$

Hipótesis alternativa    H1:      $C \neq 0$

La regla de decisión dependerá del interés y exigencia del investigador o de la investigación, recordando que el valor máximo podrá estar muy cercano a la unidad.

Ejemplo: Se desea conocer si la disposición de los trabajadores para laborar en fin de semana depende de su estado civil, o qué tanto influye éste en la decisión de participar en esos días de descanso. Los datos se presentan en la tabla siguiente:

Estado civil	¿Está dispuesto a trabajar en fin de semana?			Total
	No	Tal vez	Sí	
Casado	15	11	18	44
Soltero	24	21	17	62
Total	39	32	35	106

En este caso plantearemos la regla de decisión que evidencie, en caso de ser aceptada la hipótesis nula Ho, un alto nivel de correlación entre esas dos variables nominales.

Regla de decisión:

1. Aceptar  $H_0$  si  $C < 0.9$
2. No aceptar  $H_0$  si  $e C \geq 0.9$

Significa que en caso de ser aceptada la hipótesis nula ( $H_0$ ), la decisión de menos del 90% de los trabajadores entrevistados en su disposición por trabajar en fin de semana se encuentra influida por su estado civil.

Para obtener el coeficiente  $C$  de Pearson, estadístico de esta prueba, primero se encuentran las frecuencias esperadas según lo tratado anteriormente, obsérvelas entre paréntesis en la tabla que a continuación se presenta:

Estado civil	¿Está dispuesto a trabajar en fin de semana?			Total
	No	Tal vez	Sí	
Casado	15 <b>(16)</b>	11 <b>(13)</b>	18 <b>(15)</b>	44
Soltero	24 <b>(23)</b>	21 <b>(19)</b>	17 <b>(20)</b>	62
Total	39	32	35	106

El valor de  $X^2$  calculada se obtiene en la siguiente tabla:

Celda	Frec. observada	Frec. esperada	$(F_o - F_e)$	$(F_o - F_e)^2$	$(F_o - F_e)^2 / F_e$
1	15	16	-1	1	0.06
2	11	13	-2	4	0.31
3	18	15	3	9	0.6
4	24	23	1	1	0.04
5	21	19	2	4	0.21
6	17	20	-3	9	0.45
$X^2$ calculada					1.67

Substituyendo en el estadístico de prueba se obtiene:

$$C = \sqrt{\frac{1.67}{1.67 + 106}}$$

$C = 0.12$

Comparando el valor del coeficiente C de Pearson se toma la decisión de aceptar la hipótesis nula (  $H_0$  ) significando que existe correlación muy baja entre la disposición de trabajar en fin de semana y su estado civil, por lo que se deberán buscar otras variables que permitan determinar las causas por las que los empleados están dispuestos a laboral en días de descanso.

## GLOSARIO

- **Población:** Es un conjunto de elementos bien determinados y con características comunes, que los hace formar parte del mismo. Ejemplo: mujeres solteras que trabajan en la industria farmacéutica en el turno matutino.
  
- **Población finita:** Es aquel conjunto en el cual su número de elementos está definido o determinado. Ejemplo: los empleados que recibieron crédito hipotecario el mes pasado.
  
- **Población infinita:** Es aquel conjunto en el que el número de sus elementos es indefinido o indeterminado. Ejemplo: los trabajadores que recibirán préstamos para adquisición de automóvil durante los próximos diez años.
  
- **Muestra:** Fracción o parte de la población, esta debe ser representativa del conjunto de donde ha sido extraída; para ostentar el carácter de representatividad, la muestra deberá evidenciar tanto las coincidencias como las discrepancias existentes en la población.
  
- **Parámetro:** Es todo valor o característica que ha sido calculado, observado o extraído de la población.
  
- **Estimador:** También llamado estadístico, es todo valor o característica que ha sido calculado, observado o extraído de una muestra. El propósito del estimador es dar a conocer los parámetros sin observar a la totalidad de los elementos de la población.
  
- **Método de muestreo:** Consiste en determinar cómo serán seleccionados los elementos que habrán de integrar la muestra y se clasifica en muestreo con reemplazo y sin reemplazo.

- **Tipos de muestreo:** Se clasifica en probabilístico y no probabilístico. El primero de ellos es aquel en el que todos los elementos de la población tienen una probabilidad conocida de formar parte de la muestra. En el segundo se desconoce o es indeterminable esa probabilidad.
- **Nivel de Confianza:** De manera sencilla, se desea determinar cuántos de cada 100 elementos que se tomen como muestra, deberán poseer las características de la población. En otros términos es la probabilidad de que la estimación efectuada se acerque a la realidad.
- **Margen de error:** También llamado error muestral o nivel de precisión, es la diferencia máxima que el investigador está dispuesto a aceptar entre un estadístico y su parámetro.
- **Dispersión en la población o variabilidad:** Antes de iniciar una investigación, el equipo de proyecto deberá tener claro qué tipo de población estudiará, qué características sobresalen, cuál es la variable principal u objeto de estudio y otros aspectos que son determinantes para el involucramiento en el tema y consecuentemente para realizar diagnósticos y análisis con mayor confiabilidad y credibilidad.
- **Hipótesis:** Enunciado que afirma la relación, comportamiento, tendencia y diversos aspectos de ocurrencia entre dos o más variables de una muestra o población y está sujeto a comprobación.
- **Hipótesis de trabajo:** Es aquella que plantea el investigador o la investigación como objeto o base de su análisis y estudio; trata de dar una explicación tentativa al fenómeno que se está investigando.
- **Contraste de hipótesis:** Es un proceso estadístico que permite elegir una hipótesis de trabajo entre dos posibles y contraria una a la otra.

- **Hipótesis nula (  $H_0$  ):** Para realizar el contraste de hipótesis se supone cierta una de ellas, llamada **hipótesis nula  $H_0$** , la cual es formulada por el investigador con el propósito de ser rechazada, este enfoque resulta en ocasiones incomprensible, ya que se cuestiona ¿cómo es posible que se plantee algo que pretende rechazarse?
- **Hipótesis alternativa (  $H_1$  ):** La hipótesis alternativa se opone a la hipótesis nula y está en relación directa con el planteamiento de la hipótesis de trabajo, esto es, tienen el mismo sentido aunque planteada la primera en términos estadísticos y la última en el lenguaje de la investigación.
- **Error Tipo 1:** El Error Tipo I se comete cuando en la decisión se rechaza la Hipótesis Nula (  $H_0$  ) siendo ésta verdadera.
- **Error Tipo 11:** El Error Tipo 11 se comete cuando en la decisión se acepta la Hipótesis Nula (  $H_0$  ) siendo ésta no verdadera.
- **Nivel de significación:** Es la probabilidad de cometer el error tipo 1, esto es, rechazar la hipótesis nula (  $H_0$  ) cuando ésta es verdadera
- **Grados de libertad:** Número de posibilidades que en un conjunto de datos, se tiene para la asignación de valores, también identificado como la cantidad de variables independientes para la realización de una prueba estadística. Si los grados de libertad de un modelo tienden a incrementarse, la forma de la distribución tiende a ser normal.
- **Tabla de contingencia:** También llamada de doble entrada, es aquella que presenta matricialmente la relación entre dos variables de dos o más valores o categorías. Esa relación se observa mediante las frecuencias de incidencia entre esos elementos de las variables relacionadas. La intersección entre un renglón y una columna es llamada celda.

- **Frecuencias marginales:** Es la suma total de las frecuencias absolutas que corresponde a cada valor, intervalo o categoría de una variable que se está relacionando con otra en una tabla de contingencia o de doble entrada.
  
- **Frecuencias conjuntas:** Es la frecuencia absoluta que se observa en cada celda de una tabla de contingencia.
  
- **Frecuencias observadas:** Para el análisis de relación entre dos variables, la frecuencia observada es el número de casos que, a partir de los datos en estudio, forman parte de la intersección entre un renglón y una columna en una tabla de contingencia.
  
- **Frecuencias esperadas:** Para el análisis de relación entre dos variables, la frecuencia esperada es el número de casos que de acuerdo a la distribución marginal, la tabla de contingencia esperaba en cada celda para evidenciar la independencia entre variables.

## **PREGUNTAS FRECUENTES**

### **¿De qué tamaño debe ser una muestra para que sea representativa de la población?**

Si se pretende buscar la efectividad de una investigación, para determinar el tamaño de la muestra deberá considerarse el tamaño de la población: si la población es grande, la muestra tiende a ser relativamente pequeña; y si la población es pequeña, la muestra tenderá a ser proporcionalmente grande.

### **¿Cómo distinguir el tamaño óptimo de una muestra del diseño óptimo de la misma?**

Cuando se refiere al tamaño óptimo, lo que se busca es determinar el número de elementos que mínimamente permitirán que la estimación de los parámetros sea altamente confiable. Sin embargo, cuando se refiere el diseño óptimo de la muestra, al tamaño deberá agregársele el tipo de muestreo adecuado a las características de la población, así como el método más apropiado para la investigación.

### **¿Por qué se dice que la hipótesis nula se plantea con el propósito de ser rechazada?**

Lo que pretende el investigador es comprobar la validez de su hipótesis de trabajo y como la hipótesis nula se plantea contraria a la de trabajo; entonces, lo que se busca es encontrar evidencias para no aceptar la hipótesis nula.

### **¿Qué tipo de prueba estadística es la más recomendable en ciencias sociales?**

Como recomendación sería aventurado referir alguna prueba específica, ya que cada grupo de datos es diferente lo mismo que el planteamiento de la hipótesis. Se recomienda primeramente involucrarse ampliamente en los objetivos de la investigación y posteriormente plantear eficazmente las hipótesis de trabajo, las que traerán como consecuencia el diseño de las hipótesis nulas de donde se deriva el estadístico de prueba a utilizar.

## BIBLIOGRAFIA

Castillo Trapala, Rogelio. Los modelos reexperimentales como una Metodología de apoyo a los normalistas en ejercicio para que realicen investigación causal, ENEP ARAGÓN, UNAM, 1985.

Cochran Willam. Técnicas de muestreo, México Ed. Cecsca, 1980.

Gemmell Glass, Gene v.

Métodos estadísticos aplicados a las ciencias

sociales, México Ed. Prentice hall, 1986.

Guerrero g, Victor. Estadística básica para estudiantes de economía '1 otras ciencias sociales, México, Ed. Fondo de cultura económica, 1989.

Holguín Quiñones, Fernando. Estadística descriptiva aplicada a las ciencias sociales, México, Ed. UNAM, facultad de ciencias políticas y sociales, 1970.

Kerllnger, Frederick Nichols. Investigación del comportamiento. Técnicas y metodología, México, Ed. Trillas, (última edición)

Leach Chris. Fundamentos de estadística: enfoque no para métrico para ciencias sociales, México, Ed. Limusa, 1982.

Levin, Jack. Fundamentos de estadística en la investigación social, México, Ed. Harla, 1992.

Mateos Rivas, José. Estadística en investigación social: ejercicios resueltos, Madrid, España, Ed. Paraninfo, 1989.

Rodríguez Osuna, Jacinto. Métodos de Muestreo, Madrid, España, Ed. Centro de Investigaciones Sociológicas, 1991.

