

Unidad I

CONCEPTOS BÁSICOS

1.1 Estadística inferencial. Estadística Paramétrica y No Paramétrica

1.1.1. Estadística

Es el “conjunto de técnicas para describir grupos de datos, y para tomar decisiones en ausencia de información completa” (Young & Veldman, 1996, pág. 14) . La estadística reúne, clasifica y recuenta los hechos que tienen una determinada característica en común, para llegar a conclusiones a partir de los datos numéricos extraídos. Utiliza conjuntos de datos numéricos para obtener, a partir de ellos, *inferencias* basadas en el cálculo de probabilidades. Existen tres ramas de la estadística: descriptiva, Inferencial y proyectiva. De las cuales solo se describirá la segunda (Inferencial).

1.1.2. Inferir

Consiste en extraer un juicio o conclusión a partir de hechos, proposiciones o principios, sean generales o particulares. Es un “proceso cognitivo en el que se aplican reglas lógicas bien definidas para que, a partir de proposiciones que se consideran válidas (*premisas*) obtengan proposiciones igualmente válidas (conclusiones) (López, 2014, pág. 22). Es decir, requiere del uso de *silogismos* para lograr sus fines.

Existen dos formas de inferencia: Deducción (que va de premisas generales a conclusiones particulares, siendo de carácter determinista¹) e Inducción (“Aunque las premisas sean verdaderas, no puede asegurarse de forma determinista que no pueda llegar a encontrarse una premisa que contradiga a las ya conocidas e invalide la conclusión” (López, 2014, pág. 22), ya que la inducción contempla la posibilidad de que existan elementos o características no contempladas, por lo que es importante señalar, la importancia y necesidad de buscar la mayor cantidad de información de calidad posible, para que permita señalar con mayor precisión las conclusiones en el proceso de inducción).

1.1.3. Silogismo

Es un argumento que consta de tres proposiciones: la premisa mayor, la premisa menor y la conclusión. Esta última surge de la *deducción* de la primera (mayor) por medio de la segunda (menor). Consiste “en que en la afirmación (o negación) de un todo..., también puede afirmarse (o negarse) a cada una de las partes (Anders, s.f., pág. 3).

1.1.4. Premisa

Es la parte de un *silogismo* que afirma y niega un todo. Es un término con origen en el latín *praemissus* (**prevenido, mandado con anticipación**) (Anders, s.f.). **Se utiliza para nombrar al indicio, síntoma o conjetura que permite inferir algo** permite

¹ El determinismo señala que los fenómenos siempre se comportan de la misma forma y producen uno y sólo un resultado

sacar una conclusión. Las premisas son **aquellas proposiciones que anteceden a la conclusión**. Dicha conclusión deriva de las premisas, aunque éstas pueden ser falsas o verdaderas.

1.1.5. Conclusión

Acabar o finalizar algo.

Deducir algo después de haber

considerado Deducir algo después de haber considerado sus circunstancias (Real Academia Española, 2014)

– Premisa Nro. 1: “ <i>Todos los hombres son mortales</i> ” – Premisa Nro. 2: “ <i>Sócrates es un hombre</i> ” – Conclusión: “ <i>Sócrates es mortal</i> ”	Verdadero
– Premisa Nro. 1: “ <i>Las estrellas emiten radiación</i> ” – Premisa Nro. 2: “ <i>El sol es una estrella</i> ” – Conclusión: “ <i>El sol emite radiación</i> ”	Verdadero
– Premisa Nro. 1: “ <i>A los seres humanos les gusta el mar</i> ” – Premisa Nro. 2: “ <i>Facundo es un ser humano</i> ” – Conclusión: “ <i>A Facundo le gusta el mar</i> ”	Falso
– Premisa Nro. 1: “ <i>Todos los lunes, Fernanda se despierta a las 8 de la mañana</i> ” – Premisa Nro. 2: “ <i>Hoy, Fernanda se despertó a las 8 de la mañana</i> ” – Conclusión: “ <i>Hoy es lunes</i> ”	Falso

1.1.6. Estadística Inferencial

O *Inferencia Estadística*, es una “modalidad de inducción mediante la cual, a partir de las características de la muestra (casos particulares), llegamos a determinar las características desconocidas de la población (caso general) de la que fue extraída dicha muestra” (López, 2014, pág. 23). Es decir, se utiliza para “decidir que puede inferirse sobre grupos grandes de los que han obtenido las muestras” (Young & Veldman, 1996, pág. 14) a través de la prueba de hipótesis

La inferencia estadística se puede realizar de dos formas: a) Directamente a partir de los datos, sin la existencia de parámetros (*Inferencia No Paramétrica*); b. Utilizando parámetros (*Inferencia Paramétrica*).

Se calcula para “la estimación de los parámetros de la población y las pruebas de hipótesis” (Siegel & Castellan, 2005, pág. 23).

1.1.7. Parámetro

Es una “característica numérica de una población o de un modelo” (Everitt & Wykes, 2001, pág. 155), es decir, es “una medida estadística de la población” (López, 2011, pág. 26). El principal problema de la medición paramétrica se relaciona con el tamaño de la población, ya que entre más grande sea, mayor inversión de tiempo y recursos será necesaria para medir a todos los miembros.

Los parámetros de uso más generalizado en una población son cuatro: totales, medias, y proporciones o porcentajes (Abad & Servín, 1978, pág. 19).

1.1.8. Estadígrafo

Es “una medida de la muestra” (López, 2011, pág. 27), para más preciso es el “cálculo resumido de mediciones realizadas en una muestra para estimar un parámetro de la población” (Ritchey, 2002, pág. 37). Puntualiza Ritchey (2002, pág. 38) que en los estadígrafos² se debe reconocer que las conclusiones de una de una muestra *no son totalmente correctas*, ya que son solo *estimaciones* del parámetro y a través de los procedimientos estadísticos y la teoría de la probabilidad es posible especificar el grado de error conocido en las predicciones y el grado de confianza en una conclusión estadística.

1.1.9. Estimación Estadística

Es el “informe de una medición resumida basada en el muestreo sistemático y en mediciones precisas e informadas con grados conocidos de error y confianza” (Ritchey, 2002, pág. 38)

1.1.10. Estadística Paramétrica

Las técnicas estadísticas de estimación de parámetros, intervalos de confianza y prueba de hipótesis son, en conjunto, denominadas *Estadística Paramétrica* y se aplican básicamente a variables continuas. Se basan en especificar una forma de distribución de la variable aleatoria y de los estadísticos derivados de los datos.

La estadística paramétrica es una rama de la *estadística inferencial* que comprende los procedimientos estadísticos y de decisión basados en las distribuciones de los datos reales. Por lo que se requiere usan un número finito de parámetros. Por ejemplo, si conocemos que la altura de las personas sigue una distribución normal, pero desconocemos cuál es la media y la desviación de dicha norma. La media y la desviación típica de la desviación normal son los dos parámetros que queremos estimar.

La mayoría de procedimientos paramétricos requiere conocer la forma de distribución para las mediciones resultantes de la población estudiada. Para la inferencia paramétrica es requerida como mínimo una escala de intervalo, esto quiere decir que nuestros datos deben tener un orden y una numeración del intervalo. Es decir nuestros datos pueden estar categorizados en: menores de 20 años, de 21 a 40 años, de 41 a 60, de 61 a 80, etc. ya que hay números con los cuales realizar cálculos estadísticos. Sin embargo, datos categorizados en: niños, jóvenes, adultos y ancianos no pueden ser interpretados mediante la estadística paramétrica ya que no se puede hallar un parámetro numérico (como por ejemplo la media de edad) cuando los datos no son numéricos.

1.1.11. Estadística No Paramétrica

Rama de la estadística que estudia las pruebas y modelos estadísticos cuya distribución subyacente no se ajusta a los llamados criterios paramétricos. Su

² Nota importante, señala López (2011, p. 27) que el término *estadígrafo* es manejado por algunos autores como *estadísticos*. Lo que comúnmente origina confusiones para la comprensión y manejo de la información que se obtiene de la muestra.

distribución no puede ser definida a priori³, pues son los datos observados los que la determinan. La utilización de estos métodos se hace recomendable cuando no se puede asumir que los datos se ajusten a una distribución conocida, cuando el nivel de medida empleado no sea, como mínimo, de intervalo.

Se asume que la población de la cual la muestra es extraída es normal o aproximadamente normal. Esta propiedad es necesaria para que la prueba de hipótesis sea válida. Sin embargo, en un gran número de casos no se puede determinar la distribución original ni la distribución de los estadígrafos por lo que en realidad no tenemos parámetros a estimar. Tenemos solo distribuciones que comparar.

Cuando desconocemos totalmente que distribución siguen nuestros datos entonces deberemos aplicar primero un *test no paramétrico*, que nos ayude a conocer primero la distribución.

1.2. Muestras independientes y muestras relacionadas.

1.2.1. Muestras Independientes

Las muestras independientes son mediciones realizadas en dos conjuntos diferentes de elementos, es decir, los grupos a comparar tienen diferentes integrantes, por lo que conforman poblaciones diferentes. Si los valores de una muestra no revelan información sobre los valores de la otra muestra, entonces las muestras son independientes.

1.2.2. Muestras Relacionadas o Dependientes

Las muestras dependientes son mediciones pareadas de un conjunto de elementos, es decir, cuando los grupos a comparar tienen los mismos integrantes. Si los valores de una muestra afectan los valores de la otra muestra, entonces las muestras son dependientes, por lo que se dan medidas repetidas en el mismo grupo.

1.3. Prueba de hipótesis

Señala el Instituto Tecnológico de Chihuahua (ITCH, s.f.)

Un parámetro puede estimarse a partir de los datos contenidos en una muestra. Puede encontrarse ya sea un sólo número (*estimación puntual*) o un intervalo de valores posibles (*intervalo de confianza*). Sin embargo, muchos problemas requieren que se tome una decisión entre aceptar o rechazar una proposición sobre algún parámetro. Esta proposición recibe el nombre de *hipótesis*. Este es uno de los aspectos más útiles de la inferencia estadística, puesto que muchos tipos de problemas de toma de decisiones,

³ En general lo *a priori* se opone a lo *a posteriori*, y se refiere al paso de algo anterior a algo posterior en cualquier serie ordenada de elementos, independientemente de que esa serie tenga lugar o no en el tiempo, y de que remita a una anterioridad temporal, lógica u ontológica. En la Escolástica son llamadas *a priori* las demostraciones que van de la causa al efecto, de la esencia de un objeto a sus propiedades. Para Kant, lo *a priori* se aplica a los conocimientos cuya validez no depende de la experiencia.

pruebas o experimentos en la investigación, pueden formularse como problemas de prueba de hipótesis.

Una *hipótesis estadística* es una proposición o supuesto sobre los parámetros de una o más poblaciones.

1.3.1. Estimación puntual

Es cuando se usa un solo valor extraído de la muestra para estimar el parámetro desconocido de la población. Al valor usado se le llama *estimador*.

1.3.2. Estimación por intervalos de confianza

A veces es conveniente obtener unos límites entre los cuales se encuentre el parámetro con un cierto nivel de confianza, en este caso hablamos de estimación por intervalos

1.4. Condiciones para emplear la estadística inferencial paramétrica y no paramétrica

A decir Siegel (1980, citado por López, 2014, pág. 29) una prueba paramétrica requiere para su aplicación de cumplir las siguientes suposiciones:

- La variable es medida a nivel intervalar o de razón
- El muestreo fue aleatorio
- La muestra es suficientemente grande (Ley de los grandes números)
- Cuando se involucra una sola variable se conoce o estima su varianza
- Cuando son dos o más variables, se da la homocedasticidad entre ambas (varianzas iguales)

Agrega López (2014, pág. 29) que para una prueba no paramétrica, se considera en general la masa de datos, por lo que:

- Las variables pueden ser medidas en cualquier nivel.
- No se hace ningún tipo de suposición sobre la forma de distribución de la población (*distribución libre*)
- No importa el tamaño de la muestra, por lo que se le denomina *prueba de pequeñas muestras*.

1.5. Grados de libertad (*gl*)

Se refiere a la flexibilidad de los procedimientos estadísticos. Son el “número de oportunidades de muestreo para compensar las limitaciones, distorsiones y debilidades potenciales en los procedimientos estadísticos” (Ritchey, 2002, pág. 209).

Es el número de oportunidades en el muestreo para compensar las limitaciones, distorsiones y debilidades potenciales de los procedimientos estadísticos o el número de eventos muestrales independientes. Se calculan de diversas formas para diversos procedimientos estadísticos

1.6. Regla de decisión

	<i>Pruebas paramétricas</i>	<i>Pruebas no paramétricas</i>
<i>Nivel de medición</i>	Intervalar o superior	Desde nominal
<i>Tipo de muestreo</i>	Aleatorio	Preferentemente aleatorio
<i>Tamaño de la muestra</i>	<i>Preferentemente grande</i>	<i>Preferentemente pequeña</i>

<i>Distribución de la población</i>	Normal o conocida	Desconocido
<i>Varianza de una variable</i>	Conocida	Desconocida
<i>Varianza de dos variables</i>		Desconocida

Hipótesis Estadística

Proposición o supuesto sobre los parámetros de una o más poblaciones. Las **hipótesis** siempre son proposiciones sobre la población o distribución bajo estudio, no proposiciones sobre la muestra.

Tipos generales de hipótesis que se usan en un estudio científico

1. Hipótesis de trabajo

Sirve para intentar demostrar una relación concreta entre variables a través de un estudio científico. Estas hipótesis se verifican o se refutan por medio del método científico, por lo que en ocasiones también se conocen como *hipótesis operacionales*.

Generalmente, las hipótesis de trabajo nacen de la deducción: a partir de ciertos principios generales, el investigador asume ciertas características de un caso particular. Tienen varios subtipos: atributivas, asociativas y causales.

2. Hipótesis nula (hipótesis de no relación) (H_0)

Representa la hipótesis que se mantiene a no ser que los datos indiquen su falsedad, y puede entenderse, por tanto, en el sentido de neutra. Nunca se considera probada, aunque puede ser rechazada por los datos.

Refiere que no existe ninguna relación entre las variables que han sido objeto de investigación. No debe ser confundida con una relación negativa o inversa. Simplemente, las variables estudiadas parecen no seguir ningún patrón concreto.

Se acepta la hipótesis nula si el estudio científico da como resultado que las hipótesis de trabajo y alternativas no son observadas.

3. Hipótesis alternativa (H_1 , H_2)

Siempre se da ante la presencia de una H_0 , de manera explícita o implícita. Intentan ofrecer una respuesta a la misma pregunta que las *hipótesis de trabajo*. Explora relaciones y explicaciones distintas. De este modo es posible investigar acerca de distintas hipótesis durante el transcurso de un mismo estudio científico. Este tipo de hipótesis también puede subdividirse en atributivas, asociativas y causales.

3.1. Atributiva

La hipótesis atributiva es la que se usa para describir los hechos que ocurren entre las variables. Se usa para explicar y describir fenómenos reales y mensurables. Este tipo de hipótesis solo contiene una variable.

3.2. Asociativa

La hipótesis asociativa concreta una relación entre dos variables. En este caso, si conocemos el valor de la primera variable, podemos predecir el valor de la segunda.

3.3. Causal

La hipótesis causal establece una relación entre dos variables. Cuando una de las dos variables aumenta o disminuye, la otra sufre un aumento o disminución. Por tanto, la hipótesis causal establece una relación causa-efecto entre las variables estudiadas. Para identificar una hipótesis causal se debe establecer un vínculo de causa efecto, o relación estadística (o probabilística). También es posible verificar esta relación a través de la refutación de explicaciones alternativas. Estas hipótesis siguen la premisa: “Si X, entonces Y”.