# ANEXO 1

# Conceptos Básicos

En el presente trabajo definiremos algunos conceptos y técnicas estadísticas que se utilizaron para el desarrollo de este estudio.

**Definición de Estadística:** Kendall y Stuart afirman que: “La Estadística es la rama del método científico que trata de los datos reunidos al contar o medir las propiedades de alguna población”.

**En Resumen:** El objetivo de la Estadística es hacer inferencias con respecto a una población a partir de la información contenida en una muestra. La teoría de la estadística es una teoría de la información que trata de su cuantificación, del diseño de experimentos o procedimientos para la recopilación de datos que minimizarían el costo de una cantidad específica de información, y además del uso de esta información para hacer inferencias.

**Experimento** es el proceso por medio del cual se obtiene una observación.

El **espacio** **muestral** asociado con un experimento es el conjunto que consta de todos los posibles puntos muestrales. Un espacio muestral es denotado por Ω. A todo subconjunto A ⊂ Ω es llamado evento; Ω es un evento cierto, ∅ un evento imposible. Si ω ∈ Ω entonces ω es llamado un evento simple. Una clase Α de subconjunto Ω es llamada álgebra de subconjuntos de Ω, si satisface las siguientes propiedades:

1. Ω ∈ Α
2. Si A ∈ Α entonces Ac ∈ Α
3. Si A ∈ Α y Β ∈ Α entonces A ∪ B ∈ Α

Además si cumple que:

4. Si An ∈ Α para n = 1,2,3....., entonces 

Es llamada una σ - álgebra de subconjuntos de Ω.

Una función Ρ definida en una σ - álgebra Α es llamada medida de probabilidad en Α o simplemente probabilidad en Α si cumple que:

1. P(Α)  0

1. P(Ω ) = 1
2. Si A1,...., An son disjuntos (2 a 2 ) entonces



(dos eventos son disjuntos 2 a 2, si son mutuamente excluyentes. i.e. Ai ∩ Aj = ∅, i≠j

Una **variable aleatoria** X en un espacio de probabilidad (Ω,s,P) es una función definida en Ω tal que es un evento aleatorio para todo x∈R; i.e., X: Ω→R, es una variable aleatoria si ∈S ∀x∈R.

La **función de distribución** para una variable aleatoria X se define por:



La variable aleatoria X puede ser discreta si y sólo si le otorgan valores finito o infinito contable  tal que . La función de probabilidad de una variable aleatoria discreta es:



En este caso decimos que f es una función de probabilidad de X.

Parte de esta investigación consistirá en el estudio de las observaciones y sus características que estas poseen, así pues tenemos las siguientes definiciones:

**Población** es el conjunto total de las unidades de investigación. La información se obtiene, o se trata de obtener, de una muestra de la población para inferir características de toda la población.

**Muestra** es un subconjunto de la población que contiene unidades investigativas seleccionadas aleatoriamente.

**Parámetros** son ciertos valores que describen las características de la población, en general estos parámetros no son fáciles de encontrar, por ello se utilizan ciertas variables aleatorias para estimarlos, las cuales son denominadas **Estimadores**, en general los estimadores se basan en los datos de una muestra.

Si X es una variable aleatoria discreta , f(x) es el valor de la función de probabilidad en x y g(x) es una función de x, se define el **valor esperado** de g(x) como una transformación de E sobre g(x), tal que:



El **r-ésimo momento con respecto al origen** de la variable aleatoria X es el valor esperado de xr, representado por μ,r, así pues se tiene para X discreta:

 , r = 0,1,2......

El **r-ésimo momento con respecto a la media** de la variable aleatoria X es el valor esperado de (x-μ)r, representado por μr, se tiene para X discreta:

 , r = 0,1,2...

Entre los principales parámetros poblacionales tenemos a la **media poblacional** que se define como:

 , X una variable aleatoria discreta.

El estimador más usado de la media poblacional es la **media aritmética** que es el promedio de X1, X2,...,Xn, n observaciones de una muestra de la población:



La **Varianza Poblacional**  medida de variabilidad que se define como la media del cuadrado de las diferencias de las observaciones con respecto a la media poblacional:



Una medida de dispersión de las observaciones alrededor de la media poblacional es la **Desviación Estándar**, definiéndose como la raíz cuadrada positiva de la varianza poblacional:



Se define la **Covarianza** entre X y Y como:



Una medida de la relación lineal entre dos variables aleatorias X y Y es el **Coeficiente de Correlación**, definido por:



El **Sesgo** es otro parámetro poblacional que mide la simetría de la distribución de los datos de una población alrededor de la media.

El sesgo se calcula como:



Si el **sesgo es positivo** quiere decir que los datos se sesgan hacia la derecha, si el **sesgo es negativo** los datos se sesgan hacia la izquierda y si el **sesgo es cero**: se dice que los datos tienen distribución simétrica, entonces la media, la mediana y la moda son iguales.

La **Kurtosis** mide la picudez de la distribución de los datos de una población.

La kurtosis se calcula como: 

Si la distribución de los datos posee la forma de una distribución normal se denomina **Mesocúrtica**, si la distribución es achatada con respecto a una normal se denomina **Platicúrtica**, y **Leptocúrtica** si la distribución más puntiaguda que una normal.

**Prueba de Hipótesis**: Se utiliza para comprobar si las conjeturas con respecto a la distribución o los parámetros de una población son ciertas o no. La conjetura que se busca aceptar o rechazar se denomina **hipótesis nula (Ho)**, y la conjetura que se constratará se denomina **hipótesis alterna (H1)**.

La región o área que se utiliza para realizar inferencia se llama **región crítica de la prueba** , que representa un subconjunto Rn, tal que:



Para conocer si una prueba es rechazada o no se utiliza el **Valor p** que es el mínimo nivel de significancia de la prueba.

**Muestreo:** Es un procedimiento de investigación científica. Su función básica es determinar que parte de una realidad en estudio (población o universo) debe examinarse con la finalidad de hacer inferencias sobre dicha población. Un muestreo es la selección de una muestra a partir de una población, entendida como muestra un subconjunto, elegido de un conjunto mayor usualmente de manera aleatoria, para realizar un estudio estadístico.

Al elegir una muestra, se espera que los datos estadísticos sean proporcionales a la población. Este proceso permite ahorrar recursos, obteniendo resultados parecidos si se realizará a toda la población.

Cabe mencionar para que el muestreo sea válido y se pueda realizar un estudio fiable, debe cumplir ciertos requisitos, lo que lo convertiría en una muestra representativa.

Los términos utilizados en el muestreo son:

**Estadístico: E**s una medida usada para describir alguna característica de una muestra, tal como una media aritmética, una mediana o una desviación estándar de una muestra.

**Población Objeto:** Es un conjunto de individuos o elementos que le podemos observar, medir una característica o atributo.

**Unidades de muestreo:** La unidad de muestreo es una unidad seleccionada del marco de muestreo. Puede ser la unidad de análisis, aún cuando no es necesario.

**Unidad de Análisis:** La unidad de análisis es la unidad para la que deseamos obtener información estadística. En las encuestas de tipo usual, pueden ser personas, hogares, fincas. La unidad de análisis se denomina frecuentemente como un elemento de la población.

**Marco muestral:** La totalidad de las unidades de muestreo de donde se extraerá la muestra constituye el marco de muestreo.

**Muestra:** Conjunto de unidades o elementos de análisis sacados del marco muestral.

**Estadístico:** Los datos o medidas que se obtienen sobre una muestra y por lo tanto una estimación de los parámetros.

**Error Muestral, de estimación o standard:** Es la diferencia entre un estadístico y su parámetro correspondiente.

**Nivel de Confianza:** Probabilidad de que la estimación efectuada se ajuste a la realidad. Cualquier información que queremos recoger está distribuida según una ley de probabilidad (Gauss o Student), así llamamos nivel de confianza a la probabilidad de que el intervalo construido en torno a un estadístico capte el verdadero valor del parámetro, en un muestreo repetitivo.

La representatividad en estadística se logra con el tipo de muestreo adecuado que siempre incluye la aleatoriedad en la selección de los elementos de la población que formarán la muestra.

Una muestra es extraída aleatoriamente cuando la manera de selección es tal, que cada elemento de la población tiene la misma probabilidad de pertenecer a la muestra.

**Muestreo aleatorio simple:** Es un método de selección de n unidades en un conjunto de N de tal modo que cada una de las NCn muestras distintas tengan la misma probabilidad de ser elegidas.

**El tamaño de la muestra:** Para calcular el tamaño de la muestra de una población tenemos que tener presente si la varianza poblacional es conocida o no.

Para calcular el tamaño de la muestra es necesario estimar Image2258con un error máximo permisibleImage2259prefijado y conocida la varianza poblacional (Image2260) podemos utilizar la formula matemática:

|  |  |
| --- | --- |
| Image2261 | (1) |

que se obtiene de reconocer que Image2259es el error estándar o error máximo prefijado y está dado por la expresión Image2262para el nivel de confianza Image2263y constituye una medida de la precisión de la estimación, por lo que podemos inferir además que Image2264.

Cuando la varianza poblacional es desconocida, primero se toma una **muestra piloto**, con ella se estima la varianza poblacional (Image2260) y con este valor se halla (Image2268). El valor deImage2253obtenido será aproximadamente el valor necesario, nuevamente con ese valor de Image2253se extrae una muestra de este tamaño de la población, se le determina la varianza a esa muestra, como una segunda estimación de (Image2260) y se halla el tamaño de la muestra a aplicarse en la recolección de datos.

Se plantea esta afirmación ya que la Image2268de Image2267tiende a estabilizarse a medida que aumentaImage2253alrededor de la Image2267por lo que llegará el momento en que se encuentre el tamaño de muestra conveniente.

|  |  |
| --- | --- |
|  | (2) |

Siendo conociendo que:

Image2260es la varianza de la población respecto a determinadas variables.

Image2271es la varianza de la muestra, la cual podrá determinarse en términos de probabilidad como Image2272

Image2273es error estándar que está dado por la diferencia entre (Image2274) la media poblacional y la media muestral.

Image2275 es el error estándar al cuadrado, que sirve para determinar Image2260, por lo que Image2260=Image2276es la varianza poblacional.