###### ANEXO 2

**Conceptos Básicos**

Para tener un mayor conocimiento y entendimiento del tema a tratarse y de cómo se obtienen y el porqué de algunos supuestos es necesario tener en claro los siguientes conceptos y definiciones:

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

El proceso que se realiza para obtener una observación o medición cualquiera es lo que se denomina **Experimento**. Un experimento está asociado a varios eventos elementales o también denominados resultados posibles.

El conjunto de todos os resultados posibles de un experimento se llama **espacio muestral**, 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 llamdo 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 una σ - álgebra Α es llamada medida de probabilidad en Α o simplemente probabilidad en Α si cumple que:

1. P(Α)  0
2. P(Ω) = 1
3. 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 que el conjunto total de observaciones correspondientes a una característica de interés se llama **Población**, a la cual se le extrae un subconjunto de la población que contiene las observaciones obtenidas mediante una selección a la cual se le llama **Muestra** de la población.

Una forma estadística de describir las características de la población es por medio de ciertos valores llamados **Parámetros**, 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 ariables aleatorias X y Y es el **Coeficiente de Correlación**, definido por:



Otro parámetro poblacional es el **Sesgo** 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 atos 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 mínimo nivel de significancia de la prueba.

 **Muestreo:** El muestreo es un procedimiento de la 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( Sampling). 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 realizara a toda la población.

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

Los términos utilizados en el muestreo son:

**Estadístico:** Un estadístico es 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:** Conjunto de individuos o elementos que le podemos observar, medir una característica o atributo.

**Unidades de muestreo:** número de elementos de la población, no solapados, que se van a estudiar. Todo miembro de la población pertenecerá a una y sólo una unidad de muestreo.

**Unidades de Análisis:** Objeto o individuo del que hay que obtener información.

**Marco muestral:** lista de unidades o elementos 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.

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 formaran la muestra.

Una muestra se dice que es extraída aleatoriamente cuando la manera de selección es tal, que cada elemento de la población tiene igual oportunidad de ser seleccionado.

**Muestreo aleatorio simple:** Es aquel en que cada elemento de la población tiene la misma probabilidad de ser seleccionado para integrar la muestra.

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

Para determinar el tamaño de muestra necesario para estimar con un error máximo permisibleprefijado y conocida la varianza poblacional () podemos utilizar la formula matemática:

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

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

Si la varianza de la población es desconocida, primero se toma una pequeña muestra, que se le llama **muestra piloto**, con ella se

 estima la varianza poblacional () y con este valor se halla (). El valor deobtenido será aproximadamente el valor necesario, nuevamente con ese valor de se extrae una muestra de este tamaño de la población se le determina la varianza a esa muestra, como una segunda estimación de () y se halla el tamaño de la muestra a aplicarse en la recolección de datos.

Se puede plantear esta afirmación ya que la de tiende a estabilizarse a medida que aumentaalrededor de la por lo que llegará el momento en que se encuentre el tamaño de muestra conveniente.

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

Siendo conociendo que:

es la varianza de la población respecto a determinadas variables.

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

es error estandar que está dado por la diferencia entre () la media poblacional y la media muestral.

es el error estandar al cuadrado, que nos servirá para determinar , por lo que =es la varianza poblacional.