CAPÍTULO I

Conceptos Básicos de Estadística

CAPÍTULO I

CONCEPTOS BÁSICOS DE ESTADÍSTICA

Para realizar estudios estadísticos es necesario registrar la ocurrencia de eventos de interés a través del tiempo. Un evento está relacionado con el espacio muestral de un experimento (Hines, 1999). Existen dos tipos de eventos: determinísticos y probabilísticos. Debido a la naturaleza de esta tesis, los segundos serán los descritos.

Probabilidad se define como la frecuencia esperada de ocurrencia de un evento específico en un conjunto de posibles resultados (Hines,1999). Los resultados de los eventos probabilísticos se definen por medio de variables aleatorias que pueden tomar una serie de valores que es necesario estimarlas. Entiéndase variable aleatoria como "... una función que asigna números reales a puntos en el espacio muestral". (Tobías, 1986).

1.1 Probabilidad

El cálculo de la probabilidad de un evento determinado en el espacio muestral de un experimento se calcula de acuerdo al tipo de distribución de la población. Existen algunos principios básicos que deben de ser considerados para efectos de este cálculo.

1.1.1 Reglas básicas.

Algunas reglas básicas que nos van a ser de utilidad para el cálculo de probabilidad de un evento son las siguientes:

- Regla de Multiplicación. Se utiliza cuando se requiere conocer la probabilidad de que varios eventos independientes ocurran.

$$P(AB) = P(A)P(B)$$

- Regla de Complemento. La probabilidad de que un evento no ocurra se calcula restando 1 menos la probabilidad del evento.

$$P(A^C) = 1 - P(A)$$

1.1.2 Eventos Independientes y Dependientes

La probabilidad de que dos eventos sucedan simultáneamente (P(AB)) cuando su ocurrencia depende uno del otro se define como probabilidad condicional. Esta se puede definir como la probabilidad de que ocurra el evento A (P(A)) multiplicado por la probabilidad de que ocurra el evento B dado que sucedió el A. Es decir:

$$P(AB) = P(A)P(B/A) = P(B)P(A/B)$$

El cálculo de la probabilidad de que ocurran 2 o más eventos independientes entre sí se realiza como sigue:

$$P(ABC...N) = P(A)P(B)P(C)...P(N)$$

Gracias a esta propiedad, más adelante se podrá estimar la ecuación de la Función de Máxima Verosimilitud en forma conjunta de la distribución Tina de Baño.

1.2 Función de frecuencia.

Para lograr calcular la probabilidad de que un evento ocurra es necesario establecer previamente la frecuencia de sucesión de determinado acontecimiento. El comportamiento de sucesos varía en complejidad dependiendo del proceso que se esté estudiando.

Las distribuciones de frecuencia son fórmulas matemáticas que proveen un modelo teórico que se aproxima de manera aceptable al comportamiento de determinado tipo de datos. Estos modelos son utilizados para establecer inferencias sobre la población donde se tomo la muestra y para poder tomar decisiones con base en un estudio estadístico y probabilístico previamente realizado.

Para calcular la probabilidad de que un valor particular "x" caiga en el rango del espacio muestral "X" la función de distribución acumulativa (FDA) se enuncia como sigue:

$$F_X(x) = P_X(X \leq x)$$

La función de probabilidad o densidad se interpreta de la siguiente manera: f(x) dx es la fracción de valores de la población que ocurren en el intervalo dx (Mendenhall, 1994). En el área de confiabilidad, la mayoría de las veces, la x es sustituida por t, ya que el tiempo es la variable de interés en general.

1.3 Función de Distribución Acumulativa

En distribuciones discretas la función de densidad es la fracción de valores de la población que ocurren en el intervalo dx, es decir, es equivalente a la FDA. En distribuciones continuas la función de densidad describe el tipo de comportamiento que siguen la muestra y para obtener la FDA es necesario integrar la f(x), en el intervalo dx.

La distribución de frecuencias acumulativa (FDA) esta representada por F(x). Se relaciona con la función de probabilidad (f(x)) por medio de la siguiente igualdad:

$$F(x) = \int_{-\infty}^{t} f(x)dx$$

Donde : f(x) : Función de probabilidad o densidad.

Dado que F(x) es una probabilidad, todas las reglas y fórmulas para manipular probabilidades pueden ser utilizadas.

A continuación se enuncian las propiedades de las funciones de distribución acumulativas (Hines,1999):

1. $0 \le F_X(x) \le 1 \qquad -\infty < x < \infty$
2. $lím \xrightarrow[x \to \infty]{} F_X(x) = 1$
3. $lím \xrightarrow[x \to -\infty]{} F_X(x) = 0$
4. La función es creciente. Esto es, si $x_1 \le x_2$, entonces $F_X(x_1) \le F_X(x_2)$.

5. La función es continua desde la derecha. Esto es para todo x y $d > 0$,

$$\lim_{d \to 0} \left[F_X (x + d) - F_X (x) \right] = 0$$

1.4 Función de Máxima Verosimilitud

La Función de Máxima Verosimilitud se representa generalmente con la letra L. Es definida como la probabilidad de observar una muestra aleatoria en una población determinada. Se define de la siguiente manera:

$$L = \prod_{i=1}^{n} f(x_i) \qquad (1.1)$$

Donde: $f(x_i)$: Función de densidad ó la probabilidad de observar x_i en la muestra aleatoria.

Cabe mencionar que en sentido estricto, si tenemos una muestra aleatoria x_1, x_2, x_3,….., x_n con distribución f(x) continua; la probabilidad de que X=x se expresa como sigue:

$$P[X = x] = \int_{x}^{x} f(x)dx = 0$$

Sin embargo, coloquialmente se ha buscado la aproximación de esta probabilidad como a continuación se muestra:

$$P[X = x] = f(x)dx$$

Gracias a este convencionalismo, el Método de Máxima Verosimilitud puede calcular los estimadores insesgados (Mendenhall, 1994) de los parámetros poblacionales correspondientes. Este método considera la maximización de las funciones de densidad involucradas en el comportamiento del sistema con respecto a los parámetros de la población. Los valores, así obtenidos se denominan Estimadores de Máxima Verosimilitud (EMV).

Lógicamente, para poder aplicar este método, es necesario contar con una muestra aleatoria, es decir, con un conjunto de variables independientes e idénticamente distribuidas.

El método de máxima verosimilitud puede ser representado mediante la siguiente ecuación:

$$L\left(q^{*}\right) = \underset{q}{Max}\, L(q)$$

Debido a que el logaritmo natural es una función continua, creciente y que se maximiza en el mismo punto que la función original; es posible utilizar la siguiente igualdad para facilidad de cálculo:

$$l = \ln L \qquad (1.2)$$

Si se sustituye la ecuación 1.2 en la ecuación de la Función de Máxima Verosimilitud, ecuación 1.1, es posible escribir la siguiente ecuación:

$$l = \ln \prod_{i=1}^{n} f\left(x_i\right)$$

Una de las formas como es posible calcular los estimadores de algunas distribuciones, una vez que se obtiene la expresión "l", es derivando con respecto a cada uno de los estimadores correspondientes a cada distribución. Posteriormente igualando a cero las derivadas y se despeja el estimador deseado.

El ejemplo que se presenta a continuación permitirá entender de una mejor manera este concepto cuando los datos de la muestra siguen una distribución exponencial.

Ejemplo 1.1. Si se cuenta con una muestra aleatoria $Y_1,$ $Y_2,\ldots,Y_n$ de una distribución exponencial con una tasa de llegadas λ. Hallar el estimador de máxima verosimilitud de λ.

Solución :

$$f(x) = \lambda \exp(-\lambda x)$$

Si se reemplaza la función de densidad exponencial en la ecuación de la Función de Máxima Verosimilitud, ecuación 1.1, es posible escribir la siguiente ecuación:

$$L = \prod_{i=1}^{n} \lambda \exp(-\lambda x_i)$$

Como anteriormente se había explicado, se puede aplicar la ecuación 1.2 para facilitar el cálculo:

$$l = \ln L$$

$$= \ln\left[\prod_{i=1}^{n} \lambda \exp(-\lambda x_i)\right]$$

$$= \prod_{i=1}^{n} \ln[\lambda \exp(-\lambda x_i)]$$

$$= \prod_{i=1}^{n} [\ln \lambda - \lambda x_i]$$

$$= n \ln \lambda - \lambda \sum_{i=1}^{n} x_i$$

Debido a que el estimador a calcular es λ se procederá a derivar con respecto a este parámetro:

$$\frac{dl}{d\lambda} = \frac{n}{\lambda} - \sum_{i=1}^{n} x_i$$

A continuación se igualará la expresión anterior para maximizar la función:

$$\frac{n}{\lambda} - \sum_{i=1}^{n} x_i = 0$$

Posteriormente, se procederá a despejar λ para lograr estimarla más adelante usando una muestra aleatoria:

$$\lambda = \frac{n}{\sum_{i=1}^{n} x_i} \qquad (1.4)$$