MEDIDAS ESTADÍSTICAS, VARIABLES Y DATOS

Población y muestra. Variables y tipos. Escalas de medida. Tipos comunes de muestreo: mas, mes,…

1. Población y muestras.

Por lo general, realizamos experimentos donde reunimos un conjunto de datos que proveerán información sobre el fenómeno que esté analizando. Usando estos datos se toman conclusiones que usualmente van más allá del alcance del experimento en particular. Se generaliza desde ese experimento a la clase de todos los experimentos similares y no se puede tener la absoluta certeza de que las conclusiones sean las correctas.

Sin embargo, usando técnicas estadísticas se puede medir el grado de incertidumbre de los resultados del experimento. Estas medidas están asociadas a la eficiente selección o colección de datos que represente el experimento. En este caso se habla de muestras de observaciones sobre un fenómeno dado.

Un experimento aleatorio se define como un proceso o acción cuyo resultado no puede ser pronosticado con certidumbre y probablemente cambiaría cuan do el experimento se realiza nuevamente. La variabilidad de los resultados puede deberse a varias causas: error en las mediciones, diferentes objetos seleccionados para las pruebas, espacios disímiles, etc. La abuilidad de de modelar y analizar los resultados experimentales es el corazón de la estadística.

Uno de estos muestreos es el aleatorio simple (mas), muestro estratificado (mes), muestreo sistemático, por etapas y otros. Se recomienda consultar sobre estos tópcos si hay mayor interés en Johnson, Dallas. Métodos Multivariantes Aplicados Al Análisis De Datos. 2000. International Thomson Editores o también, Johnson, R.; Wichern, D. Applied Multivariate Statistical Analysis. 4th Ed. 1998. Prentice Hall.

Por ejemplo, algunas poblaciones tiene miembros que no existen todavía (mantenimiento preventivo) o la población como tal es demasiado grande (fuerza laboral de un estado). Los investigadores, entonces, miden sólo una parte de estas poblaciones que se le conocen como muestra.

Si se realiza inferencia acerca de la población usando la información obtenida de la muestra, entonces es importante que esta muestra sea representativa de la población. Esto puede ser posible mediante muestreo aleatorio simple, donde todas las observaciones de la población son igualmente probables de ser seleccionadas.

Para observar los datos de la muestra existe un conjunto de técnicas que permiten observar sujetos de interés y usando la información obtenida desde las observaciones de la muestra se pueden hacer afirmaciones o inferencias acerca de la población a la que la muestra se refiere. Esto se le conoce como *inferencia estadística.* Algunos de los métodos incluyen estimaciones de parámetros poblacionales (media, varianza, mediana, etc.), test de hipótesis estadística y estimaciones de funciones de densidad.

Por otro lado, la población a la que está referida la muestra o población objetivo, se refiere al conjunto completo de objetos o individuos de los cuales se necesita obtener alguna información. Deben estar bien definidos en términos de lo que constituye la población (nivel de ingreso, sector geográfico, etc.) y las características que se miden de la población (edad, altura, número de fallas, etc.)

Algunos ejemplos de poblaciones serían:

- En la fuerza laboral, la población sería toda aquella que sea económicamente activa y se mediría el nivel de ingreso, ocupación, profesión, etc.
- En presupuesto familiar, la población está referida a los hogares de una entidad y las medida estarían referidas a los ingresos de la familia y los

gastos ocasionados por cada miembro de la familia tanto en monto como en tipo.

Las variables en un experimento pueden ser discretas (ingreso mensual, edad) o continuas (peso, temperatura, fuerza, tensión de una cuerda). Las variables aleatorias discretas pueden tomarse desde conjunto de números finitos o contablemente infinitos (grandes). Mientras, una variable continua es la que puede tomar valores en un intervalo de números reales.

2. Generación de variables aleatorias. Números aleatorios uniformes.

Para ellos se emplea una función básica de MATLAB que genera una variable aleatoria uniforme. La función es **rand.** Para que reproduzca el conjunto de datos generados debe tratarse como un vector de valores aleatorios y la sintaxis aplicada en MATLAB para generar **n** números aleatorios es **rand(n).** Sin son **m** muestras de **n** números la sintaxis sería **rand(m,n)**

En este ejemplo de se muestra el uso de la función **rand**.

```
Diary on muestra
% Obtener un vector de variable aleatoria uniforme en (0,1).
x = rand(1,1000);
% Hacer un histograma.
% Se fija el ancho de la barra.
[N,X] = hist(x,15);
% Se usa la función bar para realizar el diagrama
bar(X,N,1,'w')
title('Histograma de Variables Aleatorias Uniforme)
xlabel('X')
ylabel('Freecuencia')
Diary off
```

Puede enlazar con **Colección y Procesamiento de Datos** (http://webdelprofesor.ula.ve/economia/gcolmen/programa/economia/computacion_coleccion_y_procesamiento_de_datos.pdf) para revisar algunas consideraciones básicas sobre: Análisis exploratorio de datos, variables, dato e información, gráficos estadísticos. Gráficos de sectores, diagramas de barras, histogramas de frecuencias, gráficos de caja, gráficos de tendencia, matrices

de asociación. Medidas de tendencia central: media, mediana, moda. Medidas de dispersión: rango, rango intercuartílico, varianza, desviación estándar, coeficiente de variación.

Medidas de forma: asimetría, curtosis
Medidas de asociación: covarianza, correlación, tablas cruzadas
Medidas de concentración: