



Estadística Descriptiva: Elementos Básicos

José Enrique Martín García

Universidad Politécnica de Gimialcón

(Copyright © 2016)



Población

Población estadística, también llamada universo o colectivo, es el conjunto de elementos de referencia sobre el que se realizan las observaciones.



Existen distintos tipos de poblaciones:

- **Población base:** es el grupo de personas designadas por las siguientes características. personales, geográficas o temporales, que son elegibles para participar en el estudio.
- **Población muestreada:** es la población base con criterios de viabilidad o posibilidad de realizarse el muestreo.
- **Muestra estudiada:** es el grupo de sujetos en el que se recogen los datos y se realizan las observaciones, siendo realmente un subgrupo de la población muestreada y accesible. El número de muestras que se puede obtener de una población es una o mayor de una.
- **Población diana:** es el grupo de personas a la que va proyectado dicho estudio, la clasificación característica de los mismos, lo cual lo hace modelo de estudio para el proyecto establecido.

Muestra de Población

Muestra de población, selección de un conjunto de individuos representativos de la totalidad del universo objeto de estudio, reunidos como una representación válida y de interés para la investigación de su comportamiento.

Otras definiciones altamente aceptadas, son:

- “Se llama muestra a una parte de la población a estudiar que sirve para representarla”.
Murria R. Spiegel (1991).
- “Una muestra es una colección de algunos elementos de la población, pero no de todos”.
Levin & Rubin (1996).
- “Una muestra debe ser definida en base de la población determinada, y las conclusiones que se obtengan de dicha muestra solo podrán referirse a la población en referencia”.
Cadenas (1974).



¿Qué es el muestreo?

El muestreo es el proceso de seleccionar un conjunto de individuos de una población con el fin de estudiarlos y poder caracterizar el total de la población.

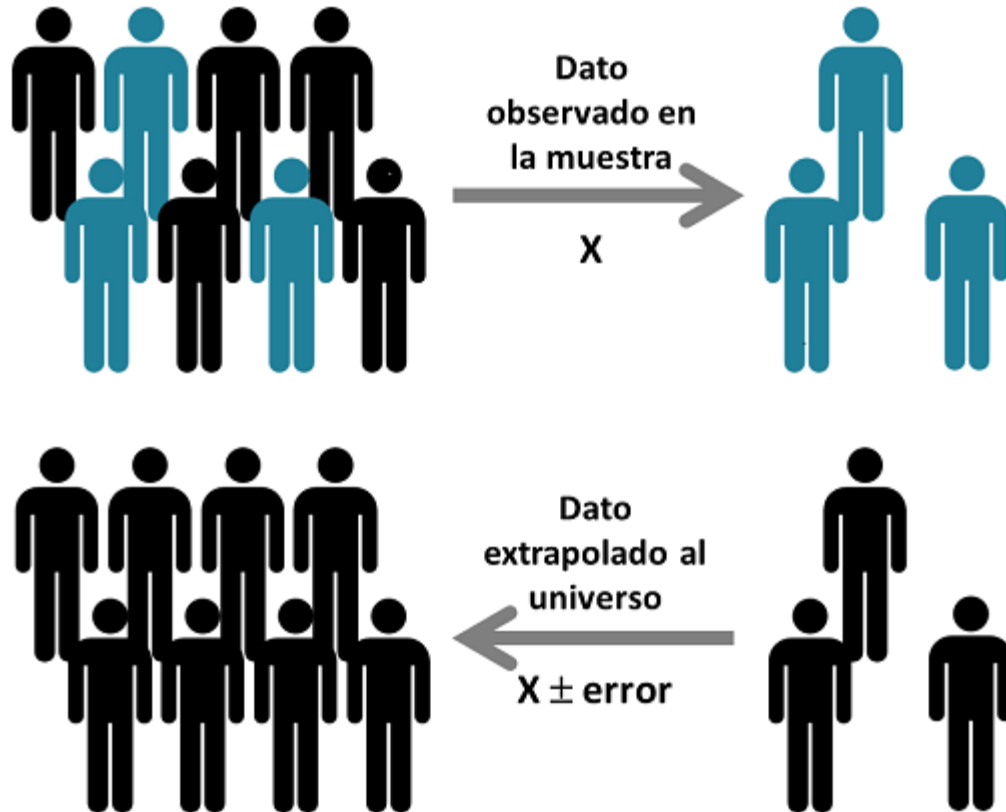
El muestreo es útil gracias a que podemos acompañarlo de un proceso inverso, que llamamos **generalización**. Es decir, para conocer un universo lo que hacemos es

- (1) extraer una muestra del mismo,
- (2) medir un dato u opinión y
- (3) proyectar en el universo el resultado observado en la muestra.

Esta proyección o extrapolación recibe el nombre de generalización de resultados.

La muestra tiene que estar protegida contra el riesgo de resultar sesgada, manipulada u orientada durante el proceso de selección, con la finalidad de proporcionar una base válida a la que se pueda aplicar la teoría de la distribución estadística

¿Qué es el muestreo?



Muestreo Aleatorio Simple

Muestreo en el que todas las muestras tienen la misma probabilidad de ser seleccionadas y en el que las unidades obtenidas a lo largo del muestreo se devuelven a la población.

Sobre una población de tamaño N podemos realizar extracciones de n elementos, pero de modo que cada vez el elemento extraído es repuesto al total del universo. De esta forma un elemento puede ser extraído varias veces

El **muestreo aleatorio con reposición** es también denominado **muestreo aleatorio simple**, que como hemos mencionado se caracteriza por que:

- Cada elemento del universo tiene la misma probabilidad de ser elegido,
- Las observaciones se realizan con reemplazamiento. De este modo, cada observación es realizada sobre el mismo universo (no disminuye con las extracciones sucesivas).
- Se garantiza la independencia entre las unidades seleccionadas.



Muestreo Aleatorio Sin Reposición

Al Muestreo aleatorio sin reposición se le denomina también muestreo irrestricto.

Cada vez que se hace una extracción, la unidad seleccionada no se devuelve al universo. Por lo tanto, no se permite que una misma unidad sea seleccionada más de una vez.

Esto hace variar la probabilidad de obtener una determinada muestra: $P(\text{obtener una determinada muestra}) = P(x_1, x_2, \dots, x_n) = 1/N \times 1/N-1 \times \dots \times 1/N-n$

Cuando la población sea muy grande, las diferencias entre estos dos tipos de muestreo van a ser mínimas.

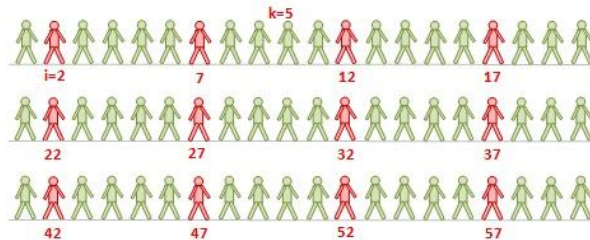
Muestreo Aleatorio Sistemático

Este procedimiento exige numerar todos los elementos de la población, pero en lugar de extraer n números aleatorios sólo se extrae uno.

Se parte de ese número aleatorio i , que es un número elegido al azar, y los elementos que integran la muestra son los que ocupa los lugares $i, i+k, i+2k, i+3k, \dots, i+(n-1)k$, es decir se toman los individuos de k en k , siendo k el resultado de dividir el tamaño de la población entre el tamaño de la muestra: $k = N/n$.

El número i que empleamos como punto de partida será un número al azar entre 1 y k .

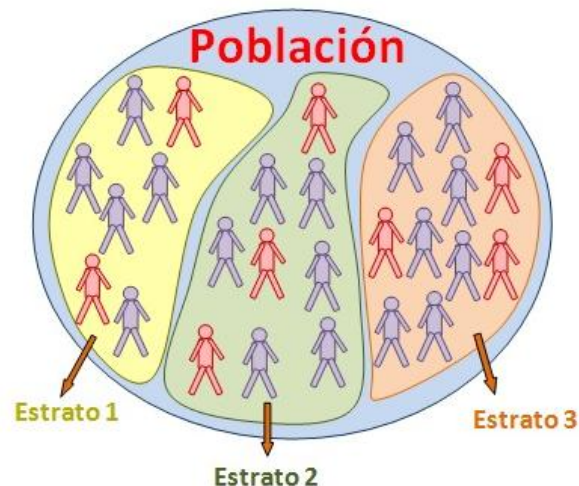
El riesgo este tipo de muestreo está en los casos en que se dan periodicidades en la población ya que al elegir a los miembros de la muestra con una periodicidad constante (k) podemos introducir una homogeneidad que no se da en la población



Muestreo Aleatorio Estratificado

Consiste en considerar categorías típicas diferentes entre sí (estratos) que poseen gran homogeneidad respecto a alguna característica. Lo que se pretende con este tipo de muestreo es asegurarse de que todos los estratos de interés estarán representados adecuadamente en la muestra.

Cada estrato funciona independientemente, pudiendo aplicarse dentro de ellos el muestreo aleatorio simple o el estratificado para elegir los elementos concretos que formarán parte de la muestra



Muestreo Aleatorio Estratificado

La distribución de la muestra en función de los diferentes estratos se denomina afijación, y puede ser de diferentes tipos:

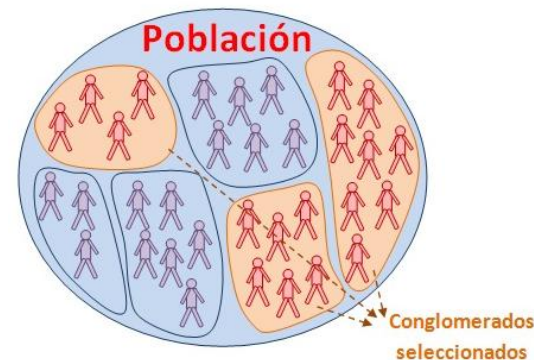
- **Afijación Simple:** A cada estrato le corresponde igual número de elementos muestrales.
- **Afijación Proporcional:** La distribución se hace de acuerdo con el peso (tamaño) de la población en cada estrato.
- **Afijación Óptima:** Se tiene en cuenta la previsible dispersión de los resultados, de modo que se considera la proporción y la desviación típica.

Muestreo Aleatorio por conglomerados

El **muestreo por conglomerados** consiste en seleccionar aleatoriamente un cierto número de conglomerados (el necesario para alcanzar el tamaño muestral establecido) y en investigar después todos los elementos pertenecientes a los conglomerados elegidos.

En el muestreo por conglomerados la unidad muestral es un grupo de elementos de la población que forman una unidad, a la que llamamos conglomerado. Las unidades hospitalarias, una caja de determinado producto, etc., son conglomerados naturales.

En otras ocasiones se pueden utilizar conglomerados no naturales como, por ejemplo, las urnas electorales. Cuando los conglomerados son áreas geográficas suele hablarse de “**muestreo por áreas**”.



Datos relacionados con el tiempo

Estadística Estática o Estructural

La estadística estática o estructural, que describe la población en un momento dado empleando datos de la actualidad (por ejemplo la tasa de nacimientos en determinado censo)

Estadística Dinámica o Evolutiva

La estadística dinámica o evolutiva, que describe como va cambiando la población en el tiempo empleando datos del pasado (por ejemplo el aumento anual en la tasa de nacimientos).





Tipos de datos

Datos cuantitativos

Son las variables que pueden medirse, cuantificarse o expresarse numéricamente y pueden ser manipulados estadísticamente. Incluyen tabulaciones de frecuencia, porcentajes, medias y promedios.

Si entre cada dos datos puede haber una infinidad de ellos, se llaman **continuos**, y si entre un dato y otro siempre hay un hueco o salto, se llaman **discretos**.

Datos cualitativos.

Son datos que no se pueden expresar numéricamente, debido a que suponen cualidades, opiniones, sentimientos entre otros, y se dividen en nominales (categorías que no mantiene relación de orden) y los jerarquizados (escalas utilizadas bajo un orden).



Variables Estadísticas

Una variable estadística es cada una de las características o cualidades que poseen los individuos de la población que estamos interesados en estudiar. Se pueden clasificar en función a la Medición o a la influencia.

Las variables cualitativas Son las variables que expresan distintas cualidades, características o modalidad

Variable cualitativa nominal: presenta modalidades no numéricas que no admiten un criterio de orden.

Variable cualitativa ordinal o variable cuasicuantitativa: La variable puede tomar distintos valores ordenados siguiendo una escala establecida, aunque no es necesario que el intervalo entre mediciones sea uniforme.



Variables Estadísticas

Una **variable cuantitativa** es la que se expresa mediante un número, por tanto se pueden realizar operaciones aritméticas con ella. Podemos distinguir dos tipos:

Variable discreta: Una variable discreta es aquella que toma valores aislados, es decir no admite valores intermedios entre dos valores específicos. Es decir, sólo puede ser expresado con números enteros.

Por ejemplo: El número de hermanos de 5 amigos: 2, 1, 0, 1, 3.

Variable continua: Una variable continua es aquella que puede tomar valores comprendidos entre dos números por lo cual tiene un número infinito de valores posibles. Es decir, puede ser expresada con números decimales o fraccionarios



Variables Dependientes e Independientes

Variables independientes son las que el investigador escoge para establecer agrupaciones en el estudio, clasificando intrínsecamente a los casos del mismo. Un tipo especial son las variables de control, que modifican al resto de las variables independientes y que de no tenerse en cuenta adecuadamente pueden alterar los resultados por medio de un sesgo.

Es aquella característica o propiedad que se supone ser la causa del fenómeno estudiado.

Variables dependientes son las variables de respuesta que se observan en el estudio y que podrían estar influenciadas por los valores de las variables independientes. La variable dependiente es el factor que es observado y medido para determinar el efecto de la variable independiente