**Resumen webinario sesión 8**

Como punto de partida es importante hablar de poblaciones. Las poblaciones que estudiamos, a veces, no existen y son construidas para que nuestros estudios las representen.

El sesgo es no tener la capacidad de describir la población que interesa.

Uno de los objetivos de la charla es discutir sobre: **¿qué es representativo? y ¿representativo de qué?**

Para dar respuesta a estas preguntas lo primero es definir la población. Esto en los casos en los que se quiere encontrar resultados representativos, es decir, aquellos que describen y miden las características de la población ya que, en la práctica, pocas veces se podrá abarcar a toda la población. Es por esto que se toman muestras. Para que una muestra pueda tomar resultados representativos, en primer lugar, debe poder dar a todos los elementos de la población la posibilidad de ser incluidos en la muestra (**posibilidad de inclusión**).

Además las muestras no son idénticas, por lo tanto, las características de las unidades que componen la muestra no son lo que las hace representativas.

La representatividad depende de las características del proceso por el que se toma la muestra.

Los procesos en los que trabajamos son complejos por lo que dar la misma posibilidad o probabilidad de inclusión a todos los elementos no es posible, ni deseable en muchos casos, pero lo que se sigue buscando es que los resultados sean representativos de la población.

Para que la probabilidad de inclusión sea la misma se usan los **procesos de muestreo aleatorio simple**. La aleatorización permite cierta “justicia” en el proceso de elección y evita los sesgos que pueden aparecer en estos procesos.

Cuando se hacen selecciones no aleatorias se está introduciendo un sesgo no manejado por leyes de probabilidad, se imposibilita la cuantificación de la variabilidad y no puede aplicarse la teoría de inferencia.

La aleatorización simple presenta desventajas pero es un proceso que permite llegar a resultados aplicables a la población. La muestra tiene la capacidad de darnos resultados que van a ser representativos. No se trata de resultados exactos, sino de estimaciones.

Sin embargo, la mayor parte de los procesos de muestreo no son simples. Un ejemplo son los **muestreos estratificados**. En este caso, primero se estratifica y luego se hace el muestreo. Para hacer una inferencia a la población total, en este tipo de muestreo, se debe hacer ajustes o “dar pesos”.

La estratificación se hace por dos razones principales:

* Se precisan resultados aplicables a grupos homogéneos y se necesita dar un tamaño de muestra específica para hacer recomendaciones. Es decir, nos asegura que un determinado grupo de interés va a tener una representación suficiente en la muestra
* Tomar en cuenta la variabilidad que conocemos de antemano

El muestreo estratificado viene determinado por los requerimientos de investigación, es decir, por las preguntas u objetivos de investigación. Ej.: puede ser interesante dentro de una población, los aspectos de género, la edad, la religión, la altitud, etc. Sin embargo, los límites en la estratificación vienen determinados por decidir qué es lo prioritario y qué es lo que se debe garantizar. Se recomienda no tener más de 4 o 5 estratos.

En estos casos, si se puede construir probabilidad de selección o pesos se puede ajustar el procesamiento de los datos de la muestra.

Algunas de las características que deben tener las muestras son:

- La posibilidad de inclusión de todos los elementos de interés. A veces esto no es posible lo que lleva a niveles de sesgo. Los sesgos son consecuencia de la realidad y, en muchos casos, son inevitables

- Los procesos de muestreo que permiten la inclusión de estratos para asegurar que los objetivos se cumplen. Hay casos donde solo interesa la información a nivel de estratos

- Reajustables a través de los pesos

Para determinar si los resultados de una muestra son representativos de la población se debe inspeccionar el proceso del muestreo y el proceso seguido para llegar a los resultados.

Por eso, el concepto de “muestra representativa” contribuye a generar confusión.

**Conglomerados**: son agrupaciones de unidades de estudio que se hacen por conveniencia, por razones prácticas. Un ejemplo son los productores y productoras de lana en el Perú. En este caso, para llegar al productor/a se parte del nivel más alto al más bajo: departamento – zona agroecológica – comunidades – hogares – productor/a.

En los niveles más altos (departamento) se suele tomar todo (todos los departamentos o regiones) por razones de negociaciones. En algunos casos (ej. para las zonas agroecológicas), se hacen conglomerados para el “trabajo físico” a través de cuadrículas artificiales para seleccionar elementos, lo que reduce la información (se elige de forma consciente).

En los niveles más bajos, los procesos aleatorios deben predominar para estimar la variabilidad de las unidades de estudio que ayudan a definir la población.

A este tipo de procesos de muestreo jerárquico se le denomina conglomerado.