

Sobre la población y muestra en investigaciones empíricas



Por José Serrano Angulo
Facultad de Ciencias de la Educación. Universidad de Málaga
Editor de sección de la revista de investigación educativa

Publicado el 13 de enero de 2017.

En las investigaciones en las Ciencias Sociales que se basan en los resultados de una muestra para generalizarlos a toda la población, se utilizan distintas pruebas estadísticas que se disponen en las oportunas aplicaciones informáticas. Con las pruebas estadísticas se puede obtener respuestas a una gran variedad de cuestiones que se plantean en las Ciencias Sociales. Estas pruebas tienen sus fundamentos en la Estadística Inferencial que determina las condiciones y los pasos a seguir en cada caso. La Estadística Inferencial no es necesaria si se cuenta con todos los datos de la población.

La generalización de los resultados de una muestra a toda la población se basa en leyes de probabilidad, por ello se debe de dejar claro cuál es la población de partida y como se elige la muestra. Aunque es bien conocida la noción de población y de muestra, no siempre se refleja en los artículos sobre investigaciones, siempre resulta más clarificador exponer brevemente el procedimiento de elección de la muestra, que denominarla con un adjetivo obviando todo el proceso de muestreo. En los artículos que se presentan trabajos con una muestra de una población, es aconsejable que la población de partida quede bien definida y se describa el modo en el que se elige la muestra.

Los conceptos de población y muestra son imprescindibles para poder entender y aplicar las distintas pruebas de la Estadística Inferencial, sin olvidar los conceptos de probabilidad. Sin entrar en muchos formalismos, se puede definir qué se entiende por población y por muestra como sigue:

- *Población*: cualquier conjunto bien definido de personas u objetos.
- *Muestra aleatoria*: muestra elegida mediante algún procedimiento de azar.

Como ya se ha señalado anteriormente, la inferencia estadística trata de generalizar los resultados obtenidos en una muestra a toda una población. Esta generalización no está exenta de errores. El error que se puede cometer se denomina error muestral.

- *Error muestral* es la diferencia entre el valor del parámetro y el valor de un estadístico, en el caso de la media se define como:

$$e = |\mu - \bar{X}|$$

Para cada muestra se tiene un error muestral distinto, en la práctica se trabaja con el error muestral máximo que se suele denotar por **E**. Este representa el error máximo que se está admitiendo bajo unas condiciones y una probabilidad determinada.

El error muestral puede ser de sesgo, o aleatorio. El error de sesgo se produce cuando la muestra no es representativa de la población, por ejemplo si se considera la población de estudiantes universitarios de una determinada Universidad y se elige una muestra de estudiantes de 1º de medicina. Aunque sea una muestra de la población de estudiantes, esta es sesgada. Así, la media de las notas de acceso a la universidad, o de la edad, o de otras variables en cuestión posiblemente se diferencie bastante de la correspondiente media de toda la población. Se puede decir que de partida esta muestra es sesgada.

Si se elige la muestra por algún procedimiento aleatorio, los posibles errores serán debidos al azar, por lo que estos errores se pueden medir con las leyes de probabilidad. Se debe elegir la muestra aleatoriamente, aun así, a priori no se tiene garantizada la representatividad de la muestra. Se pueden establecer estrategias de cómo elegir la muestra aleatoria para que la muestra sea lo más representativa posible de la población. Lo deseable es contar con una muestra aleatoria y representativa de la población. Si se elige muestras aleatorias los errores muestrales se deberán al azar y se pueden estudiar bajo las leyes de probabilidades. Hay manuales dedicados a las distintas formas de elegir la muestra aleatoria y a los cálculos de error en cada caso (Azorín, 1972, Azorín y Sánchez Crespo, 1986, Clairin, y Brion, 2001, Kish, 1979, Lohr, 2000). La representatividad de la muestra es algo que no siempre se puede garantizar y depende de que característica de la población es de interés, por ejemplo la muestra puede ser representativa según tramo de edad, la profesión, el lugar de residencia, etc., mientras más característica se consideren se puede obtener una muestra más representativa de la población, pero posiblemente se necesite para ello un tamaño inaccesible para los recursos disponible en la investigación. Por lo que se puede fijar la representatividad respecto de las variables más relevantes y que puedan marcar más las diferencias en el objetivo de la investigación.

Otra cuestión, no menos importante en la Estadística Inferencial, es el tamaño de la muestra. Así por ejemplo: si se elige una muestra de tamaño 2, seguramente la media de esa muestra y la de la población serán muy distintas. En cambio, si se elige una muestra de tamaño igual al de toda la población menos dos, la media que se obtiene en esa muestra será, si no igual, bastante próxima a la de la población. Estos casos extremos no se tiene en la práctica, pero sí cualquier otro intermedio. Esto es, la muestra con la que se trabaja tendrá un tamaño determinado, y mientras mayor sea el tamaño de la muestra, mejor precisión de los resultados. A medida que se aumenta el tamaño de la muestra se tiene errores menores (Serrano, 2010). La relación entre el tipo de error y el tamaño de muestra es el siguiente:

- Error de sesgo y tamaño: Si se tiene una muestra sesgada a mayor tamaño de la muestra se produce un error mayor.
- Error aleatorio y tamaño: Si se tiene una muestra aleatoria a mayor tamaño de la misma menor error.

Por tanto las muestras deben de ser aleatorias y cuanto más representativa de la población mejor.

En resumen: la Estadística Inferencial se basa en las leyes de probabilidad para generalizar a toda la población los resultados de una muestra, de ahí la importancia de elegir la muestra aleatoriamente, de otro modo no se tiene garantías de que los valores de significación que ofrecen los programas informáticos, o que se calculen, correspondan realmente con los resultados. Si no se dispone de recursos para elegir una muestra aleatoria de una población extensa, es conveniente determinar una población de partida más reducida en la que sí se pueda realizar el muestreo aleatorio con los recursos disponibles en la investigación.

Por lo que en los manuscritos se deberían aparecer lo siguiente:

Respecto a la población, se debe de aportar toda aquella información que se conozca y que sea relevante en la investigación, como el tamaño, las características que determinen grupos, etc. Esta información es necesaria para poder determinar a posteriori la representatividad de la muestra.

Respecto a la muestra, se debe de describir el tipo de muestreo que se ha realizado, y el método para recabar la información. Distinguiendo la muestra seleccionada inicialmente de la muestra final que responde al cuestionario. Además de describir el trabajo de campo para recaba la información.

Referencias bibliográficas:

- Azorín, F. (1972). Curso de muestreo y aplicaciones. Madrid: Aguilar
- Azorín, F. Y Sánchez Crespo, J. L. (1986). Métodos y aplicaciones del muestreo. Madrid: Alianza.
- Clairin, R. y Brion, P. (2001). Manual de Muestreo. Madrid: La Muralla.
- Kish, L. (1979). Muestreo de encuestas. México: Trilla.
- Lohr, S. L. (2000). Muestreo: Diseño y análisis. México: Thomson.
- Serrano Angulo, J. (2010). Nociones básicas sobre el análisis inferencial. En S. Nieto (Ed.) . Principios, métodos y técnicas para la investigación educativa (pp. 353-371). Madrid: Dykinson.

Cómo citar esta entrada

Serrano Angulo, J. (2017). Sobre la población y muestra en investigaciones empíricas. *Aula Magna 2.0*. [Blog]. Recuperado de: <https://cuedespyd.hypotheses.org/2353>