

SELECCIÓN DE LOS CASOS (I): Precisa Qué Habrá de Observarse

 El objetivo de esta ficha de actividad es introducir el problema de la selección de los casos una vez que se ha definido (aunque sea) provisoriamente el tipo de diseño que organizará la investigación.

 La presentación general del tema y algunos argumentos muy particulares respecto a las condiciones en que resultará más conveniente el tipo de selección no aleatoria lo encontrará en:

KING, Gary; Keohane, Robert & Verba, Sidney (2000) *El diseño de la investigación social: la inferencia en los estudios cualitativos*. Alianza Editorial. Madrid. Cap. 4 “

CORTÉS, Fernando (2004) *Selección no aleatoria y generalización*. Documento inédito. El Colegio de México, México, D.F.

 Suponiendo que ya adoptó una estrategia de diseño, es preciso ahora decidir cómo habrán de seleccionarse los casos u observaciones para su estudio.

•⇨ En términos generales, cuando corresponde es necesario precisar el universo de estudios al cual se realizará la inferencia.

✓ Es claro que los problemas de selección de casos se presentan cuando es lógicamente posible distinguir entre dos conjuntos: uno definido por comprensión (el universo) y el otro por extensión (la colección de casos investigados). Algunos tipos de investigación no admiten esta distinción básicamente porque el objeto de estudio es un acontecimiento particular. Tal es el caso de muchos estudios históricos.

 Los temas trabajados por los autores en el capítulo 4 son: i) diseño impreciso; ii) límites del muestreo aleatorio; iii) sesgo de selección y iv) selección no aleatoria.

•⇨ Un **diseño es impreciso** cuando:

✓ El número de inferencias (los parámetros a estimar) es mayor que el número de observaciones en presencia de multicolinealidad entre las variables explicativas

•⇨ **La selección aleatoria tiene restricciones.**

✓ Un punto importante aportado por estos autores es recordar que en algunas condiciones la selección aleatoria no sólo no es el tipo más eficiente para disminuir los sesgos sino que puede incrementarlos. Es decir, se muestrea aleatoriamente porque se supone que el azar distribuirá sin sesgos todas las variables perturbatorias y principales que puedan existir en el universo. Pero hay casos en que esto no puede ser posible.

✱ *Un primer y muy trivial caso es cuando el universo es reducido y por lo tanto no es posible*

SELECCIÓN DE LOS CASOS (I): Precisa Qué Habrá de Observarse

extraer una muestra estadística. Si por ejemplo, suponemos que $N=50$, la aplicación del muestreo aleatorio simple para poblaciones finitas informa que el tamaño de la muestra $n=49$.

✱ Un segundo argumento es que aún cuando el tamaño del universo pudiera ser relativamente importante, por ejemplo todas las revoluciones exitosas y fallidas que ocurrieron en América Latina desde la independencia hasta el presente, una muestra que no considerase la Revolución Cubana o la Revolución Mexicana sería teóricamente cuestionable. Evidentemente los elementos de aquel universo no tienen teóricamente la misma ponderación; no son equivalentes.

✱ En particular Cortés (2004) que existe otro caso en el que el muestreo estadístico no es susceptible de llevarse adelante y es cuando no se dispone del listado completo de las unidades de análisis.

• El **sesgo de selección** se refiere a “seleccionamos observaciones que responden a combinaciones de las variables independientes y dependientes que constatan la conclusión deseada” (King, Keohane & Verba, 2000: 139).

✓ Al prescindir de la selección aleatoria, hay que considerar de qué manera se relacionan los criterios utilizados con cada variable. Esquemáticamente, con la variable dependiente y con las variables explicativas.

✓ Esto nos conduce a una regla fundamental y evidente: **la selección tiene que permitir que se produzca al menos, alguna variación en la variable dependiente**. Existen dos formas de sesgo de selección por variable dependiente:

✱ Un primer ejemplo es aquel en que se seleccionan las observaciones tal que siempre presenten el mismo valor en la variable dependiente: investigaciones que intentan explicar los estallidos bélicos estudiando solo guerras; el comienzo de las revoluciones investigando solo estos fenómenos; las causas de la eficacia escolar analizando solo escuelas eficaces; o las pautas de asistencia a las urnas indagando solo a los que van a votar.

✱ Un segundo ejemplo de sesgo y más habitual conlleva a seleccionar en función de la variable dependiente pero restringiendo las observaciones a sólo un conjunto de los valores que ésta puede tomar. Este caso se está en presencia de un problema de **truncamiento**.

✓ “Seleccionar observaciones para un estudio mediante las categorías de la variable explicativa principal no produce problemas inferenciales porque nuestro procedimiento no predetermina el resultado del estudio, ya que no hemos restringido el grado de variación posible en la variable dependiente” (King, Keohane & Verba, 2000:147).

• **Seleccionar de manera intencionada las observaciones** implica que como mínimo, conocemos de antemano los valores de ciertas variables relevantes y descartamos la selección aleatoria.

✱ Tendremos menos posibilidades de equivocarnos si los casos se seleccionan a partir de categorías de las variables explicativas. De este modo, la propia investigación consistirá en encontrar los valores de la variable dependiente (King, Keohane & Verba, 2000:150).

✱ Como alternativa a la elección de observaciones en función de la variable explicativa, se podría partir para seleccionarlas de un abanico de valores de la variable dependiente. [...] Como

SELECCIÓN DE LOS CASOS (I): Precisa Qué Habrá de Observarse

hemos subrayado, aunque este proceso de selección pueda ayudarnos a hacer inferencias causales, carece de utilidad para hacer las descriptivas que se refieren a la variable dependiente. (King, Keohane & Verba, 2000:152).

✱ Las observaciones han de ser tan representativas como sea posible en la población sobre la que queremos generalizar.

✱ Resulta peligroso seleccionar observaciones utilizando intencionadamente tanto las variables explicativas como las dependientes, porque, al hacerlo, es fácil que, sin darnos cuenta, introduzcamos un sesgo en el resultado. (King, Keohane & Verba, 2000:153).

 Comience la aplicación de estas nociones en su proyecto a partir de las nociones de selección aleatoria y selección no aleatoria.

- ⇄ ¿En cuál esquema se ubicaría su proyecto y por qué?
- ⇄ En la exposición de King, Keohane & Verba sobre los dos grandes tipos de selección no aleatoria, se establece una diferencia de importantes consecuencias entre elegir por la variable dependiente y elegir los casos por la(s) variable(s) independientes. ¿A cuál piensa usted que se aproximaría su proyecto?

 Incorpore esta actividad a su carpeta didáctica y entréguela conjuntamente con las restantes en el momento de la evaluación.