PROTOCOLO DE OBSERVACIÓN

Luis Villarroel

Director Proyecto Centro de Estadística Aplicada - CESA

Facultad de Ciencias y Tecnología

Universidad Mayor de San Simón

E-mail lvillarr@fcyt.umss.edu.bo

Cochabamba, Diciembre 2003 - Bolivia

1. INTRODUCCIÓN

La experiencia de más de cinco años en el campo de la formación en estadística aplicada y el asesoramiento de trabajos de investigación me ha permitido identificar una importante falencia en los investigadores y principalmente alumnos que realizan proceso de planificación y de análisis de datos a través de encuestas. Por otro lado he podido constatar que la literatura en relación a estos temas es relativamente "pobre" , poco precisa y en muchos casos inexistente.

El propósito de ésta nota técnica es presentar de manera simple los principios de la elaboración de un protocolo de observación, como instrumento de base para la planificación de trabajos de investigación basados en encuestas. Por otro lado se pretende dar las bases practicas del proceso de análisis de datos, insistiendo sobre los criterios de selección de las herramientas estadísticas. La intención es llegar a quienes están confrontados con estos problemas con una visión fundamentalmente practica y de aplicación. Este documento no pretende de ninguna manera ser un manual de planificación de la investigación y menos aun un manual de análisis de datos.

2. El protocolo de observación

El protocolo de observación constituye la herramienta de planificación de procesos de generación de datos en los cuales el investigador se limita simplemente a observar, sin modificar o controlar el objeto de estudio. Diversas son las disciplinas de la ciencia que recurren a este sistema de generación de datos.

2.1. La definición de "resultados esperados"

Una revisión detallada de varios trabajos de investigación ha permitido constatar la "enorme" dificultad que existe en plantear de manera simple y clara los resultados esperados de la investigación. A menudo estos son planteados en términos de objetivos extremadamente generales y poco precisos.

Considerando que un trabajo de investigación está orientado en buscar "conocimiento" a propósito de un objeto es de importancia fundamental poder identificar de manera absolutamente precisa la NECESIDAD DE INFORMACIÓN.

En esta fase de la planificación, el investigador, debe cerrar los ojos y soñar que su trabajo ha sido concluido. ¿Qué es lo tengo como aporte?, ¿Qué conocimiento ha generado mi investigación?. El resultado de este sueño en muchos casos puede provocar que el investigador abandono el tema de su investigación. Sin embargo si el sueño ha sido "conciente" el

resultado es un listado de los "resultados esperados de la investigación".

La responsabilidad de este punto del protocolo es única y exclusivamente del especialista, el estadístico, en función de su experiencia, podría eventualmente colaborar, sin embargo corresponde al investigador tener muy clara la idea.

Al finalizar esta etapa el investigador dispone de la "primera columna" de la matriz de planificación que será presentada más adelante.

RESULTADOS ESPERADOS			
R ₁			
R ₂			
.			
R _i			
.			
R _n			

Se debe notar que el investigador, en función del tema de investigación puede plantear n resultados esperados, siendo R_i , el *iesimo* resultado.

2.2. La definición de la población y las unidades de observación

Una vez identificados los resultados esperados, el paso siguiente consiste en definir cual es el objeto de la investigación en términos de la población de estudio y las unidades de observación.

La población se define como el total de "individuos" a los cuales se interesa el investigador. El término "individuo2" es utilizado en su contexto más genérico: personas, familias, localidades, empresas, etc. La población puede ser infinita y en algunos casos finita.

Las unidades de observación constituyen precisamente los "individuos2" que serán el objeto de medición.

A menudo se encuentra en la practica de la investigación definiciones de población mal realizadas. Por ejemplo se habla de la población boliviana en trabajos orientados a estimar el ingreso promedio. El término boliviano no esta bien definido, ya que una niño de corta edad es un boliviano sin embargo no tiene ingreso, los bolivianos que no están en el territorio boliviano son o no parte de la población.

2.3. Definición de variables

Variable constituye una característica o atributo asociada a una unidad de observación. Las variables pueden ser categóricas o continuasXX

En este paso del protocolo se plantean básicamente dos problemas: ¿Cuál es el número de variables que se debe observar?, ¿cuál es la naturaleza del dato generado?. En efecto la practica de la consultancia es muy común recibir demandas que plantean ¿puede revisar mi formulario?, ¿creo que tengo muchas variables, cuales me recomienda eliminar?.

Para resolver el problema del número de variables a observar, el procedimiento es relativamente sencillo, la idea es llegar a establecer el set mínimo de variables en base a los resultados esperados. Se recomienda realizar la siguiente matriz.

RESULTADOS ESPERADOS	VARIABLES
R1	X1: Variable 1 X2: Variable 2
R2	X3: Variable 3 X1: Variable 1 X4: Variable 4
R3	X5: Variable 5

En la tabla anterior se puede observar que el investigador persigue tres resultados esperados (R1, R2, R3). Para alcanzar el primer resultado esperado se requiere evaluar dos variables, para el R2, se debe observar tres variables, sin embargo la variable X1 ya participa en el R1, finalmente el R3 requiere la observación de la variable X5. Para responder a los tres resultados esperados, es necesario observar cinco variables. La consideración de otras variables no es necesaria y si por el contrario se observan menos variables, el riesgo es no alcanzar uno o más de los resultados esperados.

La naturaleza de los datos es el segundo problema relativo al paso 3 del protocolo. Los datos son clasificados en dos categorías: datos cuantitativos, cualitativos.

Si los datos son cuantitativos ellos pueden ser de: medición (naturaleza continua), y de conteo (naturaleza discreta).

Si los datos son cualitativos ellos pueden ser: binarios, ordinales o nominales.

La definición del tipo de dato es fundamental en el paso 5 en el que se definen las herramientas de análisis de datos. Se recomienda en la tabla anterior especificar para cada una de las variables la naturaleza del dato generado cuando se evalúa la variable.

2. 4. Definición del plan de muestreo

El plan de muestreo se refiere aquellas operaciones orientadas a seleccionar individuos de la población que efectivamente serán observados constituyendo así la muestra.

El principio o regla de "oro" en el problema de la selección de la muestra es conformar un grupo de "individuos" representativos. Es importante disponer de una buena fotografía de la población de manera que el proceso de inferencia (generalización de los resultados de la muestra a la población) sea adecuado.

De manera general los planes de muestreo se clasifican en dos categorías: planes probabilísticos y no probabilísticos. Los planes probabilísticos responsabilizan al azar el proceso de selección de las unidades de la población para conformar la muestra, por el contrario los planes de muestreo no probabilísticos responsabilizan el proceso de selección al investigador. Es importante notar que en el caso de los planes de

muestreo no probabilístico supone que el investigador tenga un profundo conocimiento de la población caso contrario el proceso de selección no garantiza la representatividad de la muestra.

2.5. La determinación del tamaño de la muestra.

Este es un tema que ha menudo causa "pesadillas" en los investigadores en el entendido que es una demanda de los revisores del trabajo, quienes también a menudo desconocen el propósito de la determinación del tamaño de la muestra. Una frase común es "debes hacer aparecer una formulita que justifique tu tamaño de muestra".

El problema de la definición del tamaño de la muestra en realidad es mucho más complejo que la simple utilización de una "formulita mágica"

Muchos de mis asesorados, plantean la siguiente pregunta "quiero robarle unos minutos, solo necesito la formulita para determinar mi tamaño de muestra y luego ya no lo molesto"

De manera general, los resultados esperados de una investigación pueden tener dos propósitos:

- la estimación
- probar una hipótesis planteada.

En el caso de la estimación se requiere conocer el tamaño de la muestra de manera a lograr un estimador con una precisión deseada o admisible (d_r) a un nivel de confianza o seguridad también definido por el investigador ($1-\alpha$). Finalmente participa un tercer elemento que no puede ser fijado si no estimado por el investigador, se trata de la variabilidad asociada a la variable que estamos observando (σ). Por tanto en el caso de la estimación

$$n = f(1-\alpha, d_r, \sigma)$$

En el caso de probar un test de hipótesis la noción de precisión ya no tiene sentido, se habla de la potencia del test. De manera simple la potencia del test se define como la capacidad de la herramienta para decir la verdad. En efecto un el resultado de una prueba de hipótesis es aceptar o rechazarla. Si la herramienta acepta la hipótesis cuando en la realidad debiera rechazarse se comete un error de tipo II, y si por el contrario la herramienta rechaza una hipótesis cuando en la realidad debiera aceptarse, se comete un error de tipo I. La determinación del tamaño de la muestra, en el caso concreto de un test de hipótesis debe garantizar con un cierto nivel de confianza, que la herramienta no se equivoque.

Para poder determinar el tamaño de la muestra es necesario trabajar por resultado esperado. Previamente a la determinación del tamaño de la muestra por resultado esperado se debe realizar la selección de la herramienta en el paso siguiente.

3. El análisis de datos

El análisis de datos constituye la etapa final del protocolo de observación, de manera sencilla la intención en esta fase es lograr un aprovechamiento óptimo de los datos colectados con el propósito de alcanzar los resultados esperados.

De manera general el problema del análisis de datos plantea la dificultad de seleccionar la herramienta adecuada. La selección de la herramienta de análisis es función de cuatro criterios a considerar: a) los resultados esperados, b) el número de variables que participan en el resultado esperado, c) la naturaleza del dato y d) las condiciones de aplicación del método.

El proceso de análisis de datos se debe entender como un proceso en el cual un niño selecciona su juguete de una caja de juegos. En efecto seleccionar una herramienta está en función de factores de selección.

El proceso de análisis de datos implica una serie de etapas no necesariamente secuenciales:

- a) Análisis exploratorio de datos - EDA. El propósito en esta fase es controlar la calidad de los datos. Quiere decir identificar aquellos datos influyentes o aberrantes. Es importante ser demasiado cuidadoso con el tratamiento de los datos en ningún caso se debe proceder de manera sistemática a su eliminación. Un dato influyente debe ser objeto de una revisión responsable.
- b) Análisis descriptivo de los datos. El propósito es resumir los datos. Las herramientas de resumen de datos son diversas: tablas, gráficas y parámetros o estadígrafos.
- c) Análisis inferencial de los datos. El propósito es generalizar los resultados de la muestra a nivel poblacional. Existen dos grandes áreas: el problema de estimación en el cual se busca estimador a nivel poblacional y el problema del test de hipótesis en el que se pretende aceptar o rechazar una hipótesis planteada en el estudio.

En cada una de estas fases del análisis de datos es importante considerar el número de variables que se consideran en el resultado esperado. Si el resultado esperado hace participar una sola variable se buscan en la "caja de juguetes" herramientas de control de calidad o herramientas de resumen o herramientas de inferencia unidimensionales. Si el resultado hace participar dos variables, se buscan herramientas de control de calidad de datos, herramientas de resumen de datos o herramientas de inferencia bidimensionales. Finalmente si el resultado esperado hace participar más de dos variables, las herramientas tanto de EDA, descriptivas e inferenciales deben ser multivariantes.

El lector interesado en tener mayor información relativa a las herramientas estadísticas puede solicitar el boletín 2 (Organigrama de las principales herramientas estadísticas).

Finalmente el protocolo de observación se resume a la construcción de la matriz siguiente. Es importante notar que el investigador no debe iniciar ningún proceso de investigación si no ha definido de manera detallada la siguiente matriz.

RESULTADOS ESPERADOS	VARIABLES	METODO ESTADISTICO	TAMAÑO DE LA MUESTRA
R1	X1: Variable 1 X2: Variable 2	Método 1	Formulita 1
R2	X3: Variable 3 X1: Variable 1 X4: Variable 4	Método 2	Formulita 2
R3	X5: Variable 5	Método 3	Formulita 3

En el caso de la determinación del tamaño de la muestra es necesario ser particularmente atento al hecho siguiente: si el resultado es descriptivo no existen formulitas por el contrario si el resultado es de carácter inferencial existe una formulita o en algunos casos procedimientos más complejos de definición del tamaño de la muestra. Finalmente es importante notar que el tamaño de la muestra corresponde al valor más alto determinado por resultado esperado.

4. Conclusiones

El material presentado en esta publicación pone en evidencia de los riesgos de una mala planificación de un proceso de observación así como una propuesta de carácter metodológica que garantice una calidad científica del proceso de observación.

Finalmente el autor de esta publicación agradecerá mucho recibir comentarios, críticas.

Cochabamba, 1 de agosto 2003