

Rol de la Estadística Aplicada en Investigación Científica

Luis Antonio Villarroel Peñaranda

Centro de Estadística Aplicada (CESA)

Facultad de Ciencias y Tecnología

Universidad Mayor de San Simón

Cochabamba, Bolivia

e-mail: lvillarr@fcyt.umss.edu.bo

1. Introducción

La investigación científica, comprendida como un proceso de generación de nuevos conocimientos o la comprobación de conocimientos ya desarrollados, tiene una estrecha relación con la estadística aplicada. A menudo esta relación es mal comprendida precisamente por una mala comprensión de la estadística.

El propósito de este artículo es, por lo tanto, mostrar la relación entre la estadística aplicada y la investigación científica. Asimismo se presentan de manera más general las herramientas que ofrece la estadística aplicada a los investigadores. Finalmente se presentan algunos peligros y desafíos.

2. Estadística aplicada

En la actualidad, a causa de la expansión y diversificación de las ciencias, la literatura consagrada a la estadística es abundante y diversificada. Las aplicaciones de la estadística en las diferentes áreas de la ciencia han provocado una especie de especialización de esta disciplina: biometría, bioestadística, econometría, psicometría, sociometría, etc.

En primera instancia se pueden diferenciar aquellas referencias bibliográficas consagradas a la estadística teórica (estrecha relación con la matemática pura) y aquellas otras referencias de la estadística aplicada a una disciplina en particular. Por tanto no es extraño encontrar diversas definiciones en esta abundante literatura de la estadística dependiendo de la orientación de la referencia bibliográfica.

Dejando de lado la estadística en su contexto puramente teórico y, por el contrario, considerando su dimensión aplicada, la estadística puede ser comprendida como un

conjunto de herramientas que tienen un doble propósito:

- a) Apoyar el proceso de generación de datos: ¿cómo “juntar” datos de manera organizada?
- b) Proponer herramientas de análisis de datos: ¿cómo aprovechar los datos colectados?

En el primer caso la estadística propone dos instrumentos: el protocolo de observación y el protocolo de experimentación, según el investigador controle o no factores. En el segundo caso deben distinguirse tres fases: análisis exploratorio, análisis descriptivo y análisis inferencial.

3. Rol de la estadística en la fase de planificación de la investigación

Uno de los mayores desafíos al que estamos confrontados los profesionales dedicados a la práctica y la enseñanza de la estadística es precisamente mostrar, a los utilizadores de la estadística, la importancia y el tiempo que deben consagrar al proceso de planificación de la investigación.

Considerando que los procesos de generación de datos pueden tener dos fuentes, la simple observación y la experimentación, la estadística propone dos herramientas: el protocolo de observación y el protocolo de experimentación.

3.1. El protocolo de observación

El protocolo de observación implica desarrollar un “documento de trabajo” que contemple mínimamente los siguientes puntos:

- a) *Definición de resultados esperados.* El investigador debe identificar, de manera muy precisa, cuáles los resultados esperados de la investigación. A menudo éstos son formulados como preguntas. En la mayor parte de estos trabajos se plantean objetivos que tienen un carácter muy general y no permiten identificar cuál es el aporte concreto de la investigación. Existe una regla práctica para lograr la formulación de resultados esperados: “se ha concluido el trabajo de investigación, ¿qué es lo que se ha logrado?”. Si no es posible visualizar el trabajo finalizado, no es posible tener claros los resultados esperados de la investigación. Esta fase de la planificación debe concluir con una lista de resultados, que denominaremos R_i .
- b) *Definición de la población y unidades de observación.* El resultado de esta etapa es la delimitación del objeto de investigación. ¿Quiénes son los “individuos” que serán observados y cuántos son?
- c) *Definición de las variables.* Se debe identificar para cada resultado esperado (R_i), cuáles son las variables (X_{ij}) que se necesita observar para lograr el resultado. En este punto se plantean dos problemas: cuántas variables deben observarse, con qué pre-

cisión deben observarse las variables. El número de variables a observar está en función de los i resultados y el número de variables j por resultado esperado. Según el dominio de las variables observadas y la precisión con la que se evalúan éstas, distinguimos los datos cuantitativos (de naturaleza continua o discreta) y los datos cualitativos (binarios, ordinales o nominales). Es importante destacar que cada una de las variables debe ser calificada en términos del tipo de dato que se genera. En el caso de las variables cualitativas se recomienda elaborar un manual de codificación con el propósito de facilitar el tratamiento informático de los datos.

- d) *Definición del método de colecta de datos.* Es importante en el protocolo definir cuál o cuáles serán los métodos de colecta de datos. Diversas son las opciones: entrevistas estructuradas o no estructuradas, encuestas, talleres, grupos de expertos, etc.
- e) *Definición del plan de muestreo.* En muchos casos, no es posible observar al conjunto de la población, por falta de recursos, sean éstos financieros o de tiempo, situación en la que el investigador debe proceder a la observación de una parte de la población. El plan de muestreo es el conjunto de operaciones que permiten seleccionar las unidades de la población que serán efectivamente observados. Los planes de muestreo pueden ser de dos tipos: probabilísticos o no probabilísticos. En el caso de los planes probabilísticos se “responsabiliza” al azar de la selección de las unidades de formación. Por el contrario, en el caso de los planes de muestreo no probabilístico, la responsabilidad de la selección de unidades de observación recae sobre el juicio del investigador. Es evidente que en este segundo caso se requiere un buen conocimiento de la población para garantizar la representatividad de la muestra.
- f) *Definición del tamaño de la muestra.* La definición del tamaño de la muestra es un problema relativamente complejo que en ningún caso se resume a la aplicación de una fórmula mágica. Es importante hacer notar que la definición del tamaño de la muestra tiene sentido únicamente en aquellos resultados que requieren de la aplicación de métodos estadísticos (inferencia estadística). Si el problema es de estimación, el tamaño de la muestra debe garantizar una precisión predeterminada. Por otro lado, si el problema es relativo a un test estadístico, el tamaño de la muestra debe garantizar una potencia (probabilidad de rechazar la hipótesis cuando ésta es falsa en la realidad) preestablecida. El tamaño de la muestra debe ser obtenido para cada resultado R_i , en función del número de variables asociado al resultado y la naturaleza del dato. Por tanto, en un trabajo de investigación pueden existir varias fórmulas. En algunos casos la solución no es exacta y se requiere de procedimientos mucho más complejos.
- g) *Definición de los métodos estadísticos.* Finalmente, se debe proceder a la definición de métodos estadísticos por resultado esperado R_i . Esto implica que al análisis de los datos no puede ser realizado por un analista de datos que desconoce los resultados esperados y las variables que participan por cada R_i .

3.2. El protocolo de experimentación

El protocolo de experimentación implica la definición de los siguientes puntos:

- a) *Definición de resultados esperados y condiciones de aplicación.* Como en el caso del protocolo de observación, anteriormente presentado, el primer paso del protocolo de experimentación es la formulación de los resultados R_i . Por otro lado es importante definir las condiciones en las cuales se realiza la experimentación: ensayo en condiciones artificiales o reales, ensayo preliminar, principal o de confirmación.
- b) *Definición de factores.* Se denomina factor a una variable de naturaleza cualitativa que tiene incidencia sobre otra variable denominada de observación. Los factores tienen asociados niveles o variantes. Los factores pueden ser nominales u ordinales.
- c) *Definición de las unidades de experimentación.* Las unidades experimentales son aquellas que reciben un nivel o variante del factor y son el objeto de medición. Son las unidades en las que se evalúan los efectos del factor.
- d) *Definición de observaciones.* Como en el caso del protocolo de observación, se deben definir las variables que van a ser observadas en cada unidad experimental, así como variables accesorias. La variable en la que se evalúa el efecto del factor se denomina variable de respuesta.
- e) *Definición del diseño experimental.* El diseño experimental corresponde a la manera como se asignan los niveles del factor a las unidades experimentales. Existen diversos diseños: completamente aleatorio, en bloques, cuadrado latino, parcelas divididas, criss-cross y cros-over.
- f) *Definición del modelo matemático.* En el campo de los diseños experimentales se han propuesto modelos matemáticos asociados a diversos diseños. A menudo éstos son modelos lineales.

4. El análisis de los datos

En el análisis de los datos se deben considerar tres etapas:

- a) *Análisis exploratorio de datos.* En la literatura en inglés se conoce como herramientas EDA (*Exploring Data Analysis*). El propósito de las herramientas EDA es controlar la calidad de los datos.
- b) *Análisis descriptivo de los datos.* El propósito de las herramientas de la estadística descriptiva es resumir los datos.
- c) *Análisis inferencial de los datos.* El propósito de la herramienta de inferencia estadística es generalizar los resultados obtenidos en la muestra hacia la población.

La selección de una herramienta de control de calidad o de estadística descriptiva debe basarse en el análisis de dos criterios de selección:

- a) Número de variables que participan en el resultado esperado R_i .
- b) Naturaleza de los datos.

En el caso de las herramientas de inferencia estadística, se debe considerar:

- a) Número de variables que participan en el resultado esperado R_i .
- b) Naturaleza de los datos.
- c) Condiciones de aplicación de los métodos estadísticos.

La estadística descriptiva debe ser comprendida como el conjunto de herramientas que ayudan a resumir los datos. La inferencia estadística, por su lado, debe ser entendida como el conjunto de herramientas que permiten generalizar los datos obtenidos en una muestra hacia la población. Es importante notar que entre la estadística descriptiva y la inferencia estadística existe un "puente" que reúne todas las bases teóricas que permiten precisamente realizar la inferencia estadística. Este "puente" reúne: la teoría de probabilidades, la noción de variable aleatoria, la noción de distribuciones de probabilidad, las principales distribuciones teóricas discretas y continuas y bases de la inferencia estadística.

Finalmente es importante distinguir entre: herramientas unidimensionales, cuando se trabaja con una sola variable, bidimensionales, cuando se trabaja con dos variables simultáneamente, y multidimensionales, cuando se trabaja con más de dos variables al mismo tiempo, tanto para las herramientas de carácter descriptivo como para las herramientas de carácter inferencial.

5. Algunos desafíos y peligros

5.1. Actitud de los investigadores

La actitud individualista y el aislamiento de algunos investigadores representan un gran peligro para alcanzar resultados con calidad científica. Actualmente, el desarrollo de la ciencia implica el concurso de equipos multidisciplinarios, con alto espíritu de equipo y alta calidad académica. La integración de los especialistas en estadística en los programas de investigación debe ser comprendida en un sentido productivo. El rol del estadístico en la investigación no representa un trabajo de segundo orden. Se debe valorar el trabajo del verdadero Estadístico-Consultor.

El trabajo interdisciplinario implica una actitud "humilde" tanto del investigador como del estadístico. A menudo, los problemas que se presentan al especialista en estadística le son desconocidos. El investigador debe comprender esta situación y comunicarle de manera sencilla el marco teórico de su investigación. Por su parte, el estadístico debe comprender que las bases teóricas de los investigadores en el campo de la estadística pueden ser muy débiles y que su rol es hacerles comprender, de manera sencilla, sus debilidades. Esta práctica no siempre es evidente, lo que puede provocar situaciones muy difíciles. El investigador debe reconocer el aporte del estadístico y éste debe comprender que sus herramientas no tienen mayor sentido si no son aplicadas a favor de la investigación.

5.2. Reducida cantidad de profesionales formados en estadística aplicada

De manera general en Bolivia, como en el caso de muchos otros países el número de profesionales formados en el área de estadística, capaces de establecer un equilibrio entre la teoría y la práctica es muy reducido. Se deben realizar esfuerzos institucionales importantes para mejorar la formación de base de los estadísticos en materia de aplicación y asesoramiento, así como sus bases teóricas. En el caso de los "no estadísticos", una oferta de formación permanente o continua, formal o informal, debe ser una prioridad de los responsables de institutos de investigación, proyectos, etc.

5.3. Rol de la informática en la estadística

Con el adelanto de las ciencias de la computación y el desarrollo de software estadístico cada vez más accesible a los utilizadores de la estadística, se plantea un gran desafío para el profesional estadístico: poner en evidencia el peligro que representa el uso de estas herramientas sin bases teóricas sólidas en estadística.

A excepción del paquete SAS, la mayor parte de los programas estadísticos comerciales, SPSS, MINITAB, STATISTICS, etc., operan con menús desplegables muy fáciles de manejar y, por tanto, accesible a cualquier tipo de usuario.

En el campo de la formación es importante desarrollar una propuesta imaginativa para poder integrar un paquete informático cuando se enseña la estadística. A menudo se escucha: "*para qué hacer la teoría cuando el paquete puede hacer todo*". Considero que los responsables de la formación deben tener propuestas que eviten precisamente este tipo de comentarios. El profesor de la materia debe organizar las prácticas de computador para reforzar y desarrollar operaciones que no pueden ser presentadas utilizando los medios didácticos tradicionales (a menudo una simple pizarra). El rol de un paquete en el proceso de formación debe ir orientado a desarrollar las habilidades de aplicabilidad, síntesis y análisis en los participantes.

A manera de conclusión cabe resaltar el rol central de la estadística aplicada en el proceso de investigación científica, tanto en el nivel de planificación como en el nivel de aprovechamiento de los datos, y la urgente necesidad de fomentar un trabajo conjunto entre el investigador y el estadístico.