

PREGUNTAS Y CRÍTICAS FRECUENTEMENTE FORMULADAS SOBRE LAS ESTADÍSTICAS DE LA COMISIÓN DE LA VERDAD Y RECONCILIACIÓN DEL PERÚ

Daniel Manrique y David Sulmont

Documento elaborado entre el 2006 y 2007, con textos escritos desde el 2004

V1.1 Julio 2007

Sobre este documento

La presente es una compilación de preguntas y cuestionamientos frecuentemente formulados sobre las estadísticas publicadas por la Comisión de la Verdad y Reconciliación en su Informe Final.

Las preguntas y cuestionamientos recopilados han sido obtenidos investigando en los archivos de algunas listas de discusión en Internet y artículos periodísticos. En muchos casos han sido reformuladas, eliminando mucho del contenido no pertinente, haciéndolas más directas, e incluso en muchos casos más serias de lo que originalmente eran. Creemos haber sido justos con las preguntas y críticas planteadas, no simplificándolas para que sean fáciles de refutar ni tergiversado su sentido.

Nuestro propósito al hacer público este documento se limita a responder a estos cuestionamientos. No entramos a discutir aspectos "de fondo" relacionados con la interpretación de estos resultados ni de las conclusiones de la CVR. Aunque creemos que la discusión importante gira en torno a esos temas, también creemos que es importante aclarar algunos malentendidos acerca del estudio estadístico de la CVR debido a que las conclusiones derivadas éste son capitales para una adecuada comprensión del conflicto armado interno que sufrió el Perú entre 1980 y 2000.

¿Por qué no se ha respondido antes a estos cuestionamientos?

Varias de las preguntas y cuestionamientos que aparecen en esta lista han sido respondidas cuando han sido formuladas, tanto en presentaciones públicas como en algunos escritos circulados en Internet desde el 2003 hasta la fecha. Algunos cuestionamientos, sobre todo los de carácter más técnico, han sido contestados de forma más bien parcial y cualitativamente, aunque a nuestro juicio las críticas más severas han sido adecuadamente respondidas. Adicionalmente, en un gran número de casos la simple pero atenta lectura de lo ya publicado bastaba para absolver las preguntas planteadas.

Es importante entender que la CVR dejó de existir en agosto del 2003. Las personas que conformaron su equipo técnico tienen la convicción de haber realizado un buen trabajo y siempre estarán dispuestas a defenderlo pero, como el resto del mundo, también tienen otras ocupaciones y no pueden salir al frente de cada acusación cada vez que se plantea, menos aún si, como suelen estarlo, vienen cargadas de insultos, burlas y otros comentarios impertinentes. Finalmente, muchos de los cuestionamientos *ya respondidos* son levantados una y otra vez y muchos de ellos son francamente deleznable.

Este documento ha tenido una prolongada elaboración que cubre cerca de tres años, con largos periodos durmiendo en nuestros discos duros, a la espera de juntar el tiempo y la energía necesarios para poder armar una versión

medianamente coherente y publicarla. Esperamos que esta compilación sea de utilidad y que todos nos podamos dedicar a asuntos más constructivos.

¿Qué tipo de preguntas y cuestionamientos se han formulado?

Podemos clasificar las preguntas y cuestionamientos que hemos recibido en cuatro grandes grupos:

- Generalidades: Preguntas sobre la necesidad de hacer una estimación y en general sobre cuál era el estado del conocimiento antes de haber efectuado el estudio.
- Sentido común sobre cuestiones técnicas: Preguntas sobre cuestiones técnicas que se puede hacer el lector no especializado.
- Cuestiones técnicas: Preguntas que se hace el lector técnico.
- Interpretación: Preguntas sobre la interpretación de los resultados presentados.
- Misceláneas: Las que no caen en ninguna de las categorías previas.

Hemos organizado este documento siguiendo esta clasificación para facilitar su consulta.

1. Prolegómenos

1.1) ¿De qué datos realmente disponía la comisión?

La CVR llevó a cabo un recojo masivo de testimonios a nivel nacional y logró recopilar 16,886. Del análisis de éstos se logró individualizar a 23,140 personas muertas o desaparecidas por su nombre y apellido *paterno*, y a 18,397 de éstas además por el apellido materno. El número de víctimas de otras violaciones no fatales (esto es, contando a las víctimas que sobrevivieron) es aún mayor, aunque los conteos son complicados debido a que la misma persona puede sufrir más de una violación en diferentes momentos y lugares.

Adicionalmente, la CVR contó con información procesada por otras instituciones desde la década de 1980: el Ministerio Público; las organizaciones de la Coordinadora Nacional de Derechos Humanos; COMISEDH; el Centro de Desarrollo Agropecuario; la Cruz Roja Internacional y la Defensoría del Pueblo. Combinando los registros existentes en estas fuentes con los datos recogidos directamente por la CVR y eliminando las duplicaciones sumamos 24,692 muertos y desaparecidos *con nombres y apellidos completos* (paterno y materno). Es importante no confundir estos 24,692 individuos identificados con los 23,140 que se mencionaron en el párrafo anterior: la similitud en los números es una coincidencia fortuita que puede inducir fácilmente a asociaciones erróneas⁴.

Si sumáramos los registros de personas total y parcialmente identificadas en todas las fuentes y en las listas entregadas por las fuerzas del orden, eliminando las duplicaciones, se sobrepasaría fácilmente los 35000. Sin embargo este número está sujeto a errores no fácilmente controlables y por ello nunca ha sido sostenido.

1.2) Ok., entonces se identificaron 24,692 muertos y desaparecidos. ¿Por qué la CVR sostiene que fueron cerca de 70000?

Cuando en el contexto del trabajo de la CVR se hace mención a la cifra 24,692, se entiende que son los registros de personas muertas o desaparecidas, con nombres y apellidos completos obtenidas del cruce de las bases de datos existentes. Lamentablemente (y creemos que esto es algo que nadie discute) ningún proyecto de este tipo ha compilado una lista completa. Esto se hace evidente si notamos que todas estas listas incluyen individuos que las otras no ¿acaso hay alguna razón para creer que un proyecto adicional no adolecería del mismo problema?

La pregunta que se desprende inmediatamente es: ¿cuántos peruanos murieron en el conflicto armado interno y *no fueron registrados*? Para responderla la CVR aplicó un conjunto de técnicas estadísticas llamadas "Estimación de Múltiples Sistemas" (EMS) usando los datos de los que disponía y llegó a la conclusión de que la cifra total de peruanos muertos y desaparecidos se ubica alrededor de los 70000. El detalle técnico de esta proyección se encuentra en el anexo 3 del Informe final.

2. Preguntas y críticas de sentido común sobre cuestiones técnicas

2.1) Cuál es la diferencia entre la “captura-recaptura” y la “estimación de múltiples sistemas”

Básicamente, ninguna. El nombre utilizado depende del dominio de aplicación. El método es llamado captura-recaptura en las ciencias biológicas, donde es típicamente usado para estimar la abundancia de especies animales. El nombre se deriva del procedimiento empleado para obtener los datos: capturar el animal, dejarlo libre y recapturarlo. El método clásico de la captura-recaptura se basa en estas dos etapas (captura y recaptura) y cuando se efectúan más de dos etapas se habla de “captura-recaptura múltiple”.

La estimación de múltiples sistemas (en adelante EMS) se refiere al trabajo con listas estáticas generadas bajo distintos procedimientos, y es el nombre con el que se conoce al procedimiento en epidemiología, demografía y ciencias sociales en general. En este caso, la “recaptura” es equivalente al cruce de información entre las listas.

La diferencia de nombres se puede volver relevante debido a que distintos dominios de aplicación presentan distintos retos. En las ciencias biológicas, por ejemplo, es muy relevante el problema de animales perdiendo sus marcas entre capturas y recapturas, así como el problema de animales aprendiendo a evitar las trampas después de la primera captura. En el caso de la EMS, lo que resulta ser más importante son los procedimientos bajo los cuales se han construido las listas y la calidad del cruce entre ellas.

2.2) ¿Qué evaluaciones de consistencia se ha hecho a las estimaciones? Y en general, ¿Qué evidencias hay de que las estimaciones son razonables?

Podemos dividir estas evaluaciones en evaluaciones de consistencia interna y consistencia externa. El primer grupo consiste en pruebas de que las conclusiones que se desprenden de los procedimientos de estimación no se contradicen entre sí; mientras que el segundo, que estos resultados no contradicen aquel conocimiento *bien establecido* que se puede obtener por otros medios. Una tercera categoría, que

en sentido estricto se solapa con las anteriores, podría pensarse en términos de indicios razonables de sentido común.

Dentro del primer grupo, coherencia interna, una de las mejores pruebas se encuentra detallada en el Informe Final: Se ha estimado las mismas cantidades usando diferentes estimadores y *diferentes datos* y se ha obtenido básicamente los mismos resultados. Este resultado tiene mucha fuerza, debido a que de alguna forma nos está indicando que el esquema de post estratificación (ver sección técnica para más detalle) está realmente controlando gran parte de la heterogeneidad y que el resto está siendo controlado razonablemente bien por el modelado estadístico.

En el segundo grupo podemos citar el ejercicio de análisis demográfico realizado para el departamento de Ayacucho, que también se encuentra en el Informe Final. Utilizando los datos de los censos de 1981 y 1993, en conjunto con las estadísticas demográficas producidas por el INEI, un simple ejercicio de balance demográfico nos muestra que la población esperada en este departamento para el año 1993 es mayor que la observada en el censo. Una posible explicación (y desde nuestro punto de vista la más razonable) es la mortalidad excesiva causada por el conflicto, que es consistente con este análisis.

Otro indicio poderoso, que no ha sido publicado, es la similitud entre la estimación calculada por la CVR con el método EMS para Lima Metropolitana y el Callao (1,931), y el conteo simple realizado por DESCO para esta misma localidad[8], basado exclusivamente en información periodística (2,105). Puesto que la cobertura dada por la prensa al conflicto ha sido particularmente minuciosa *sólo* en la ciudad de Lima, éste es un fuerte indicio de que el método ha producido resultados bastante razonables.

2.3) Yo hice el experimento de “captura-recaptura” en casa usando frijolitos de diferentes colores y obtuve los siguientes resultados: de un total de 100 frijolitos, marqué 5, los mezclé con los restantes y cogí 59 al azar. De esos, sólo uno estaba marcado. La “estimación” resultante sería $N=5*59/1 = 295$. ¡El método no sirve!

A nadie se le ocurriría hacer una encuesta con una muestra de 2 individuos. Si, por ejemplo, los dos individuos seleccionados respondieran que votaron por Humala en las elecciones del 2006, ¿significaría esto que el 100% del electorado votó por Humala? Definitivamente, no.

El experimento al que se hace mención tiene el mismo defecto que nuestra hipotética encuesta. Las muestras tomadas son simplemente muy pequeñas (en particular el error fue tomar sólo 5 frijolitos en la primera muestra) y la estimación resultante tiene un error estadístico demasiado alto. En este caso, el *error estándar* es nada menos que de 152(!) El enunciado correcto, basado en un intervalo de confianza al 95%, sería: “El total de frijolitos debe estar en algún punto entre 0 y 599”, información que es cierta, pero inútil! (Sin embargo, siendo justos con el método y el experimento, hay que reconocer que algo de información nos proporciona: al menos sabemos que es altamente improbable que haya más de 600 frijolitos! En la vida real, éste dato podría ser útil si ésta fuera la primera aproximación al problema).

2.4) ¿Por qué se ha dado la cifra exacta de 69280 muertos y desaparecidos? Lo correcto hubiera sido proporcionar un intervalo de confianza o un margen de error.

Totalmente de acuerdo. Y es lo que se ha hecho. En el anexo en donde se detallan las estimaciones del número de víctimas, todas las estimaciones van acompañadas de su respectivo intervalo de confianza al 95%.

La confusión proviene del hecho de que en las conclusiones generales de la CVR, en el tomo XIII de su Informe Final, se haga mención a la cifra exacta ("estimación puntual", en jerga técnica), cosa que aunque es debidamente explicada en un pie de página, no deja de inducir a error. Un fraseo más adecuado sería algo como "alrededor de 70 mil víctimas".

3. Cuestiones técnicas

3.1) El método de captura-recaptura implica que no ha habido migraciones en la población bajo estudio. Todos sabemos que durante los años de conflicto hubo masivas olas migratorias (¡que no fueron aleatorias, para colmo!). El método está mal aplicado porque se está violando un supuesto fundamental.

Respuesta corta: Los muertos no migran.

Este cuestionamiento y otros similares reflejan una confusión que aparece frecuentemente en las críticas: ¿cuál es nuestra unidad de análisis? La respuesta es simple pero frecuentemente malentendida. En esta aplicación no estamos estudiando directamente las *muertes y desapariciones*, como una lectura superficial sugeriría, sino *reportes* de esas muertes y desapariciones. La diferencia es sutil, pero importante. Lo que buscamos es justamente estimar cuántos individuos *no han sido reportados*.

En otros contextos el tema de las migraciones puede ser muy importante, debido a que si un individuo saliera del área de estudio después de haber aparecido en una muestra, su probabilidad de captura en posteriores oportunidades sería nula. En el caso de la EMS aplicada al caso que nos ocupa, los reportes son siempre retrospectivos y proporcionados por informantes: no tenemos que "buscar" al individuo sino contar con alguien que sepa de su muerte. Dicho de otra manera: si una persona murió en una circunstancia determinada, no puede migrar y morir de nuevo en otra por más que los testigos hayan migrado a la China. Visto así, el problema de las migraciones, aleatorias o no, simplemente desaparece.

3.2) ¿Por qué se utilizó el método jackknife para la estimación de los intervalos de confianza de los estimadores?

El problema está en que, debido a escasez de información para las víctimas de Sendero Luminoso, las estimaciones por agente responsable están basadas en diferencias de estimaciones para agentes combinados. En concreto: $[SLU] = [SLU+EST] - [EST]$ (para detalles sobre las siglas ver el anexo del Informe Final de la CVR)

Si bien en la literatura se pueden encontrar fórmulas desarrolladas para el cómputo del error asintótico marginal de las estimaciones parciales ($[SLU+EST]$ y $[EST]$),

para estimar el error de la diferencia, sería necesario conocer la distribución conjunta, o al menos las covarianzas entre ambos estimadores. Si bien podría hacerse el desarrollo asintótico y derivar un estimador, el uso de métodos intensivos en cómputo es más sencillo.

La elección de esta técnica en particular sobre otras (bootstrap no paramétrico, por ejemplo), fue una cuestión más bien de conveniencia: en esta aplicación el jackknife requiere de un menor tiempo de cómputo.

3.3) Ok, pero la aplicación es incorrecta. Se ha construido los intervalos de confianza suponiendo normalidad y estamos ante un problema de conteo, no de estimación de una tendencia central; la población no es normal y el teorema del límite central no es aplicable.

Los estimadores utilizados son de máxima verosimilitud condicional bajo muestreo multinomial. Como es bien sabido, los estimadores de este tipo son en general asintóticamente normales y se dispone de un Teorema Central del Límite (TCL). Para este problema en particular, las distribuciones asintóticas fueron derivadas por L. Sanathanan en 1972 [3].

El cómputo llevado a cabo no es un "conteo" sino la estimación de un parámetro. Aunque se trate de parámetro discreto, el estimador usado no deja de ser un estimador de máxima verosimilitud como cualquier otro y, *dentro de ciertas restricciones*, podemos aplicar el TCL. Además ¿quién dice que el TCL no se puede aplicar a conteos? Las típicas aproximaciones normales a las distribuciones de Poisson y binomial son aplicaciones del TCL a variables aleatorias discretas que pueden ser interpretadas como conteos.

Dicho esto, estamos de acuerdo con que sería adecuado intentar utilizar otra estrategia para computar los intervalos de confianza. Algunos estudios utilizando métodos bayesianos (véase p.e. [6]) muestran que en muchos casos las distribuciones posteriores del parámetro N (población total) pueden ser más bien asimétricas, con una cola pesada hacia valores superiores. Jana Asher en un ejercicio no publicado computó intervalos de confianza asimétricos usando el método del perfil de verosimilitud (profile likelihood) de Cormack [7] para las víctimas de agentes del estado y encontró que, interesantemente, la mayoría estaban completamente contenidos en los intervalos simétricos derivados de la aproximación normal. Si la situación fuera similar para el resto de estimaciones, los intervalos de confianza construidos por la CVR serían más bien conservadores, pero sería interesante comprobarlo.

3.4) uhm... ¿pero por qué no computaron intervalos de confianza basados en percentiles de la distribución empírica derivada del remuestreo efectuado en vez de usar esta cuestionable suposición de normalidad?

Se está confundiendo la técnica jackknife con bootstrap. Jackknife nos brinda una estimación del error estándar del estimador, pero no nos ayuda a construir una distribución empírica como en el caso del bootstrap [2]. Sería muy interesante y probablemente muy informativo construir un intervalo empírico usando bootstrap.

3.5) La especificación del modelo utilizado hace que la estimación de las víctimas no reportadas sea proporcional a la

cantidad de víctimas reportadas. Esto produce una sobreestimación en el caso de tener listas exhaustivas y una subestimación en caso de tener listas muy incompletas. El efecto neto sería una sobreestimación.

Las estimaciones de los totales que se hacen en la EMS dependen de la *estructura* de reporte y solapamiento entre las fuentes. Ni siquiera en el caso de la estimación de dos sistemas se puede predecir, sólo a partir de qué tan exhaustiva es una lista, cuánto será la estimación; mucho menos en una estimación con tres sistemas, donde además de por medio hay modelado y estratificación.

Es difícil entender cómo se puede haber llegado a una conclusión tan disparatada. Haciendo un esfuerzo de interpretación podemos adivinar que confusiones como ésta provienen de creer que los modelos usados especifican una relación entre el gran total (N) y variables predictoras proporcionales a la "exhaustividad de captura" o algo similar, al estilo de un típico modelo de regresión, como los que se usan rutinariamente en la econometría. Aclaremos desde ahora que esto no es cierto.

Quienes hayan trabajado en análisis de clasificaciones cruzadas usando modelos log lineales recordarán que normalmente N es considerado una constante dada en la función de verosimilitud, no un parámetro a estimar. La EMS es diferente: en este caso N es considerado un parámetro del modelo (de hecho, el único que nos interesa). Afortunadamente, gracias a los resultados derivados por L. Sanathanan y S. Fienberg en la década del 70, el uso de la técnica de estimación por máxima verosimilitud condicional nos permite aplicar directamente las herramientas del análisis log lineal a nuestro problema, simplificando el análisis y los cálculos involucrados.

3.6) El método supone homogeneidad en la población e independencia entre las listas. Ninguno de estos supuestos es razonable.

Tanto la falta de homogeneidad (igualdad en las probabilidades de captura a lo largo de toda la población de estudio) como la posible dependencia de los procesos de captura para cada lista (la pertenencia de un elemento a una lista afecta su probabilidad de pertenencia a otra) son, en efecto, factores que dificultan las estimaciones. Los autores del estudio han reconocido estos temas y han incluido una discusión sobre sus efectos y la forma en la que los han enfrentado.

Para empezar una aclaración: homogeneidad e independencia **no son supuestos de la EMS**. Cuando en el texto del anexo que detalla el procedimiento de estimación se hace referencia a estas limitaciones, se hace en el contexto de la captura-recaptura clásica, que es equivalente a la versión de 2 listas de los métodos utilizados. Acto seguido, se presentan los métodos utilizados para enfrentar el problema:

"Si existieran listas adicionales de miembros de la población que se está estudiando, los problemas arriba mencionados de dependencia o heterogeneidad podrían resolverse mediante la elaboración de modelos, que es justamente lo que hacemos más adelante".

La otra estrategia está presentada algunas líneas más abajo:

"Sin embargo, a menudo las transgresiones de los supuestos se vuelven demasiado complicadas como para que la estimación de múltiples sistemas las pueda incorporar [...] Cuando ninguno de los modelos es apropiado, una opción es estratificar a la población, utilizando para ello alguna variable que sea capaz de reflejar las razones que se encuentran detrás de la heterogeneidad de las probabilidades de que un determinado individuo sea capturado en alguna listas (véase, por ejemplo, Sekar y Deming 1949)".

Entonces, valga la primera aclaración: ¡Nadie discute la existencia de heterogeneidad en la población y la posibilidad de dependencia entre las listas (muy probablemente inducida por dicha heterogeneidad individual)! Presentar estos fenómenos como un "descubrimiento" que "invalida" los resultados del estudio no es otra cosa que repetir lo que ya ha sido dicho en el propio informe, sacándolo de contexto. Por poner una analogía, esta crítica es equivalente a postular la invalidez de la metodología de encuestas de opinión porque "las preferencias de los sectores A, B, C, D y E son claramente heterogéneas". Como en el caso de las encuestas, el asunto que nos debe ocupar no es la existencia de fenómenos que nadie discute, *sino la forma en la que han sido enfrentados*.

La primera estrategia es el modelado. El aparato de los modelos loglineales, de uso corriente en el análisis de la estructura de probabilidad en las tablas de contingencia derivadas de conteos en clasificaciones cruzadas, nos brinda una poderosa herramienta para controlar las posibles distorsiones originadas por la estructura de dependencia en la población. Partiendo del hecho indiscutido de que los procesos de recolección de datos han sido distintos para la creación de cada lista, podemos notar que la forma en que esta heterogeneidad se manifiesta es originando dependencias entre las celdas de las tablas de contingencia resultantes (este hecho sorprendentemente juega a nuestro favor en la estimación). El aparato del análisis log-lineal justamente nos permite modelar estas dependencias, postulando diversas relaciones estructurales de probabilidades de captura (desiguales), *poniendo a prueba esas hipótesis* y eligiendo modelos adecuados, a fin de *encontrar* un buen conjunto de supuestos que nos permitan superar las distorsiones inducidas por la dependencia (para más detalle véase [5] y [4]).

Sin embargo, por más poderosa que sea esta metodología, existe un límite en cuanto a la complejidad de la estructura de heterogeneidades que estos modelos pueden manejar, por lo que puede hacerse necesaria la aplicación de otra estrategia para controlar su efecto.

La segunda estrategia empleada es la estratificación. La estratificación consiste simplemente en dividir nuestra muestra utilizando una característica que pueda aislar alguna de las fuentes de heterogeneidad. De esta forma se realiza una estimación para cada estrato (potencialmente con diferentes modelos), con la esperanza de que cada uno de estos estratos sea al menos más homogéneo que toda la muestra [5]. En el estudio de la CVR la primera variable elegida fue la localidad geográfica que, como ha sido resaltado numerosas veces por los críticos, puede ser una enorme fuente de heterogeneidad de captura. La segunda variable, que ha recibido mucha menor atención, es el agente perpetrador de la muerte. Esta última es muy importante dado que en el caso de la lista construida por la CVR las proporciones de víctimas de Sendero Luminoso y del Estado Peruano son radicalmente distintas a las encontradas en los otros proyectos de recolección de información.

El proceso de estratificación geográfica no fue un capricho al azar. Se siguió un procedimiento sistemático de subdivisión del territorio nacional en regiones cada

vez más pequeñas, pero que compartían continuidad geográfica y, de acuerdo al mejor conocimiento disponible, dinámicas similares. La subdivisión geográfica permitió obtener modelos que pasaran las pruebas de ajuste, que (quien quiera puede tomarse el trabajo de verificar esto) la agregación total de los datos, o subdivisiones demasiado grandes, precisamente por heterogeneidad, no permiten.

Por supuesto, ningún método estadístico hace magia y, dependiendo de qué tan severos sean estos problemas, el método usado podría arrojar resultados distorsionados. Sin embargo creemos haber tomado precauciones bastante razonables (que han sido el grueso del trabajo) que los críticos simplemente escogen ignorar o calificar de "jerga inútil" cuando formulan este tipo de objeciones.

3.7) El método supone múltiples repeticiones del experimento para lograr la estimación. La CVR sólo tuvo tres listas. ¿Cómo esperan haber obtenido un resultado razonable?

Existe una severa confusión en esta afirmación. El equivalente del "experimento" en la EMS no es la creación de una lista sino *cada reporte*. En este sentido, el "experimento" se ha repetido al menos 30000 veces, un número bastante grande.

3.8) La CVR ha usado un modelo con una serie de supuestos y nos ha presentado un número sin hacer un análisis de cuánto cambiarían los resultados si es que cambian los supuestos.

Esto es falso. Una parte importante y particularmente laboriosa del estudio fue el proceso de selección de modelos. Como señala acertadamente la crítica, un modelo estadístico expresa un conjunto de supuestos sobre los datos, que deben ser puestos a prueba de alguna forma. La selección de modelos consiste justamente en poner a prueba esos supuestos y elegir aquellos que son consistentes con lo que se observa (los datos). En este caso, los diferentes modelos representan diferentes supuestos sobre la estructura de dependencia de las probabilidades de reporte en las listas.

En general en los procesos de selección de modelos se trata de balancear dos cualidades opuestas: simplicidad y ajuste. De los muchos criterios para la selección de modelos, el estudio de la CVR se basó en la minimización del cociente X^2/df , donde X^2 es el estadístico chi-cuadrado y df representa los grados de libertad del modelo (véase [1] para más detalle). Adicionalmente, aunque esto no está documentado en el informe, se hizo el análisis usando el estadístico G^2 (logaritmo de la razón de verosimilitudes; véase [5]) en vez de X^2 y el BIC (criterio de información bayesiano). En todos estos casos los resultados fueron muy similares.

En esta misma línea también se ha criticado el hecho de que la especificación de los modelos usados (log lineales jerárquicos) "sólo" permita representar 7 diferentes estructuras de probabilidad. Sin embargo se suele perder de vista que se ha realizado una estimación diferente (con un modelo diferente) para cada uno de los 59 estratos definidos. Efectivamente esto significa que el espacio de posibles modelos para la estimación completa es nada menos que de

$$7^{59} = 72574551534231909331741171093173785967490646405143$$

posibles modelos (!), de los cuales se escogió uno. (esto, por supuesto, no es prueba de que los modelos sean adecuados -para eso se realizó la selección de modelos- aunque nos ayuda a poner la crítica en contexto!).

¿Se podría haber aplicado otros modelos? Por supuesto. De hecho en los últimos años se ha desarrollado una multitud de nuevos modelos, cada vez más sofisticados, que representan supuestos cada vez más sofisticados (interesantemente muchas de estas innovaciones, bajo algunas restricciones, son equivalentes a los modelos log lineales clásicos). Sin embargo, la metodología empleada, combinada con el proceso de estratificación, ha probado a lo largo de los últimos treinta años ser de por sí bastante consistente y robusta en situaciones similares (sobre todo en epidemiología) y no creemos que haya motivos serios para creer que los resultados arrojados sean incorrectos.

4. Preguntas de interpretación

4.1) Para que la cifra de 69280 sea cierta, 9.48 personas tendrían que haber sido asesinadas por día. Eso no es posible.

Para que la cifra de 35000 muertos y desaparecidos sea cierta, tendrían que haber muerto 4.8 personas diarias; ¿Nos dice algo esta afirmación?. Las 70000 víctimas estimadas suponen aproximadamente el 0.3% de la población peruana de la década de los 80; si fueran 35000, eso equivale aproximadamente a 0.15%...; podemos hacer miles de ejercicios aritméticos de este tipo, pero ¿qué significan estos datos?. En realidad una cifra por sí misma, fuera de su contexto es difícil de explicar. Cuantitativamente, 35 mil o 70 mil representan una pequeñísima fracción de la población peruana. En el terremoto del Callejón de Huaylas en 1971 se calcula que perecieron cerca de 60 mil personas (en un solo día!), ¿eso hace que las muertes por violencia política sean más o menos graves?, ¿tiene sentido comparar ambos acontecimientos?

Para interpretar adecuadamente los datos cuantitativos, especialmente los de este tipo, debemos tomar en cuenta el contexto histórico, social y geográfico en el que sucedieron los hechos. Existen numerosos indicios y evidencias recogidas antes, durante y después del trabajo de la CVR que dan cuenta que en el caso de las zonas más afectadas por la violencia política, lo que ocurrió fue realmente una hecatombe social y política de dimensiones tales que incluso tienen repercusiones en la dinámica demográfica de esas localidades, tal y como lo sugieren los datos del censo del 93 si los comparamos con los del 81, incluyendo el efecto de la migración y del desplazamiento interno. El problema, y es justamente una de las conclusiones de la CVR, es que las zonas donde esto ocurrió tienen poca relevancia política, social y económica para el conjunto del país, por lo tanto lo que sucedía ahí era menos conocido y menos comprendido por la mayoría de la población peruana. Una vez que analizamos con detenimiento los fenómenos históricos, la interpretación de los múltiples indicadores que podemos observar adquiere sentido y coherencia pues da cuenta de un proceso y de una historia (o de varias historias) que requieren ser reconstruidas y contadas. El mandato encomendado por el Estado peruano a la CVR fue emprender justamente ese tipo de tarea.

4.2) La CVR ha inflado las cifras de forma tal que la responsabilidad de Sendero Luminoso / los agentes del Estado se vea aumentada / disminuida [Nota: hay versiones de este comentario para todas las combinaciones]

Lo curioso de este argumento es que ha sido esgrimido en ambos sentidos (por diferentes personas). Para algunos las cifras de la CVR se han "inflado" para evitar decir que el Estado fue el principal responsable de las víctimas fatales ocurridas durante el conflicto. Para otros se "inflan" para que el Estado aparezca teniendo mayor responsabilidad de la que en realidad tuvo. Es cierto que el conflicto armado interno peruano es singular en relación a otros conflictos en la región (Guatemala, Chile, Argentina), en los que el Estado aparece claramente como el responsable de más del 90% de las víctimas fatales. La peculiaridad del caso peruano y de la letalidad de la acción de Sendero Luminoso está asociada a la política y estrategia de "guerra popular prolongada" que esta organización subversiva intentó poner en práctica en nuestro país y que se reconstruye con detalle en el Informe Final de la CVR.

En el caso de la actuación de los Agentes del Estado, en el Informe Final de la CVR se proporcionan múltiples argumentos y evidencias que dan cuenta que en determinados periodos y lugares hubo una práctica masiva y sistemática de violación de derechos humanos que se saldó en un número importante de víctimas fatales, especialmente a inicios del conflicto (1983-1984) en el departamento de Ayacucho.

Las estimaciones cuantitativas respecto de quién provocó más víctimas son producto de un análisis científico desprovisto de prejuicios políticos. Las conclusiones de ese análisis tienen consecuencias políticas e históricas que nuestra sociedad debe aún procesar si queremos cerrar definitivamente las heridas de este conflicto y mirar con confianza el futuro. Ignorar o tergiversar la historia porque no nos gusta la imagen que pinta de nuestro país no resuelve esos problemas.

4.3) ¿Por qué recién en el estudio de la CVR “aparecieron” todas las víctimas de Sendero Luminoso?

Las bases de datos que documentan la identidad de las víctimas y que fueron elaboradas por diversas instituciones (organizaciones de derechos humanos, la Defensoría del Pueblo, entre otras) antes de la existencia de la CVR contienen sobre todo denuncias de familiares y de víctimas de violaciones a los derechos humanos cometidas por agentes del Estado. El Estado peruano tenía la obligación legal de combatir la amenaza a la seguridad pública y a la democracia que significó la acción de los grupos subversivos, pero también estaba obligado a combatir esa amenaza usando las herramientas y procedimientos del Estado de derecho, que era justamente lo que los grupos subversivos buscaban destruir. La misión institucional de las organizaciones de derechos humanos era promover y vigilar que el Estado peruano cumpla con sus compromisos y la legislación que garantiza los derechos fundamentales de las personas, incluso (o más bien, sobre todo) en contextos de violencia política, en el marco de esa misión recibieron y documentaron numerosas denuncias de violaciones cometidas por agentes del Estado.

Las fuerzas del orden (la policía, las fuerzas armadas) tenían la responsabilidad de documentar adecuadamente las denuncias y a las personas afectadas por el accionar de los grupos subversivos. El problema es que mucha de esa documentación no se sistematizó adecuadamente ni fue de dominio público. Cuando la CVR solicitó al Ministerio de Defensa información al respecto recibió una escasa respuesta. Incluso en el caso de los efectivos de las fuerzas del orden que ofrendaron su vida o su salud en el cumplimiento de su deber, el Estado no fue capaz de elaborar un registro exhaustivo y detallado (problema que se dio particularmente en el caso de la Policía Nacional).

Si bien muchas de las víctimas provocadas por los grupos subversivos, sobre todo Sendero Luminoso, fueron conocidas por medio de la prensa, ninguna institución previa a la CVR hizo una documentación exhaustiva de esos casos destinada a identificar con precisión quiénes eran esas víctimas. La CVR tuvo como mandato registrar e identificar a las víctimas del conflicto armado interno, independientemente del agente perpetrador, y diseñó estrategias específicamente destinadas a registrar y documentar la mayor cantidad posible de casos, usando todas las fuentes disponibles y promoviendo el recojo de testimonios a nivel nacional. Es por ello que recién en esa oportunidad se construye un sistema que logra registrar a miles de víctimas que hasta entonces no habían sido apropiadamente documentadas, especialmente aquellas causadas por el accionar de Sendero Luminoso.

5. Misceláneas

5.1) El método estadístico usado por la CVR sólo es aplicable en las ciencias naturales para estimar poblaciones animales pero no a entornos tan sofisticados como al caso de muertes en el conflicto peruano.

Esta crítica tiene dos partes. La primera (la exclusividad del método en ciencias naturales) implica que las técnicas estadísticas desarrolladas para resolver ciertos problemas no pueden aplicarse en contextos diferentes a los que le dieron origen. En realidad muchas innovaciones técnicas "viajan" entre diferentes "realidades" si es que demuestran su utilidad para comprender y resolver los problemas que surgen en ellas. En sus orígenes las matemáticas se desarrollaron para contar objetos (por ejemplo personas, animales, o cantidades de granos producto de las cosechas) y fueron complejizándose para resolver problemas relacionados con la necesidad de construir edificaciones cada vez más grandes u observar el movimiento de los astros (ésto último asociado además a la magia y la religión). Sin embargo a nadie se le ocurriría decir que la aritmética, el álgebra o la geometría sólo sirven en la contabilidad, la arquitectura, la ingeniería civil, la astrología o la astronomía.

En el caso de la estimación de múltiples sistemas, si bien el método fue originalmente pensado para resolver un problema relacionado con la industria pesquera escandinava, en la actualidad es ampliamente empleado (con modificaciones) en la epidemiología y demografía para estimar poblaciones humanas. Estas aplicaciones van desde las correcciones censales (donde se usa la estimación de sólo dos sistemas) hasta la estimación de poblaciones elusivas y de difícil enumeración, como las personas sin hogar, drogadictos, o personas que padecen una enfermedad. Incluso se ha utilizado en experimentos para detectar errores de programación en los códigos del software de navegación del trasbordador espacial.

La segunda parte de la crítica podría ser razonable, pero se exagera en pensar que la realidad peruana y el conflicto armado que vivió el país, por más compleja que sea, es la única realidad compleja que existe en el mundo. La estimación de muertes en el conflicto peruano implica, decididamente, tomar en cuenta un entorno bastante peculiar y difícil, al igual que lo fue en el caso de Guatemala, Kósovo o Timor Oriental, otros lugares donde se usó la metodología EMS para enfrentar problemas similares. Problemas análogamente complejos son abordados por epidemiólogos, demógrafos, biólogos, sociólogos, usando técnicas similares. El reto consiste en adaptar este tipo de herramientas para que sean útiles ante el problema que tenemos en frente.

5.2) Los “expertos” de la CVR han usado modelos “log lineales” y estimación “jackknife” para decorar su trabajo y mostrar su sofisticación econométrica.

Las razones del uso de estos métodos han sido expuestas en algunas de las respuestas anteriores (véase sección 3). Como se explicó en (5.1), la EMS es una técnica que se puede situar entre la demografía matemática, la bioestadística y la epidemiología, que fue desarrollada para enfrentar problemas muy distintos de los que se encuentran usualmente en la econometría.

5.3) La captura-recaptura es un método decimonónico (sic) con un error superior al 50%

El error de cualquier estimación estadística depende de varios factores (los principales son el tamaño de la muestra y la eficiencia del estimador) y, de acuerdo a ellos, puede ser muy variable. La captura-recaptura puede tener errores mayores que el 50% (como en el ejemplo de los frijolitos en alguna de las preguntas anteriores) o incluso podría tener error 0% (imagine que ha logrado censar a toda su población). En todo caso, lo importante no es que el error exista, sino que seamos capaces de cuantificarlo. En el caso de la estimación de la CVR el error estimado ha sido del 12%.

5.4) La CVR ha asumido que las proporciones de responsabilidades de las muertes y desapariciones reportadas a ella son las correctas y ha proyectado esa proporción a todas las fuentes para obtener su estimación.

Parte de las preocupaciones de la CVR era justamente que no se sabía qué tan correcta era esta proporción. Ésa es una de las razones que llevaron a hacer este estudio estadístico. El método empleado para la estimación (EMS) dista mucho de simplemente extrapolar las proporciones.

Sin embargo, un hecho innegable es que la CVR recibió una inesperada cantidad de reportes de víctimas de SL, de las que no se había tenido noticia antes.

5.5) La CVR ha hecho su estimación basándose en la diferencia poblacional entre los censos de 1981 y 1993. La supuesta gente faltante podría haber muerto por otras razones o podría haber migrado.

De nuevo, el método utilizado para la estimación no tiene nada que ver con lo que se afirma en esta crítica. Sin embargo, es cierto que se realizó un ejercicio de balance demográfico como el descrito, sólo para efectos de evaluar la consistencia de las estimaciones. Este ejercicio fue realizado sólo para el departamento de Ayacucho y *sí se tomó en cuenta el efecto de la migración*, usando los saldos y tasas de migración estimados con los datos del censo del 93. Los resultados de este ejercicio son consistentes con los resultados de la estimación usando EMS, mostrando una población "faltante" algo mayor a la mortandad estimada usando la EMS.

5.6) La CVR ha contado reportes de víctimas y no ha tomado en cuenta que algunos de esos reportes podrían estar repetidos. Esto lleva a una sobreestimación.

La estimación realizada requirió como preparación básica de los datos justamente la depuración de la base de datos, no sólo al interior de la información de la CVR sino con la información proporcionada por diversas organizaciones que venían sistematizando su información desde la década de los 80. En ese proceso se detectó que no sólo "algunos" sino varios miles de registros estaban, en efecto, duplicados. Ésa es justamente la información sobre la que se construyeron las estimaciones, que distan mucho de ser simples conteos.

5.7) Yo conozco el distrito de Chungui, en Ayacucho. Según CEDAP, en este distrito se produjeron 982 muertes y desapariciones atribuibles al conflicto, sobre una población de 4,400 habitantes. La CVR contó 619, de los cuales 258 ya estaban en el registro de CEDAP y 361 eran nuevos. Aplicando el "método de Ball", la proyección sería $982 * 619 / 258 = 2,356$. Eso es el 54% de la población de Chungui, lo que está lejos de la verdad. Si bien Chungui fue el lugar más afectado por la violencia, de acuerdo a lo que yo conozco, allí debieron morir alrededor del 30% de sus habitantes, pero no el 54%.

Lo interesante de esta crítica es que plantea un caso concreto, con datos y números concretos (que nosotros mismos hemos proporcionado), aunque utilizados de una forma bastante torpe.

Todos los números son correctos y, debido quizás a las características propias del distrito de Chungui y de los procesos de levantamiento de información llevados a cabo, esta estimación ingenua, con sólo dos sistemas, es bastante similar a la de tres sistemas (hay que notar que rara vez éste es el caso). Sin embargo, si nos atenemos a los datos mostrados podemos encontrar fácilmente la fuente de la falacia: la base de población tomada es nada menos que la del censo del 93, o sea, icuando toda esta gente *ya estaba muerta* y muchos habían migrado! Bajo este torpe criterio metodológico, esta "fracción" de la población total que fue asesinada podría haber superado el 100% y *no hubiera habido contradicción alguna*.

Usando un criterio un poco más razonable, con los datos del censo de 1981 (justo antes de la escalada descontrolada de la violencia, en 1984), tenemos que la población del distrito era de 8257 habitantes (el doble de lo que se afirma en la crítica) y por lo tanto la fracción de la población total del distrito que fue asesinada durante el conflicto armado interno fue del 28%, una cifra espeluznante, pero no irreal. Es más, es menor que el 30% $((982+619-258)/4400)$, producto del conteo simple y la mala aplicación del censo, que al crítico no le causaba extrañeza alguna.

Vale la pena notar que este ejercicio nos está mostrando que de acuerdo a los censos la población chunguina se redujo en casi un 50% entre 1981 y 1993, hecho que ilustra dramáticamente las dimensiones catastróficas del conflicto en esa región. Esta crítica en realidad brinda elementos de apoyo al trabajo realizado.

5.8) La CVR ha creído en este resultado porque fue presentado por un gringo y los peruanos aceptan cualquier cosa que digan los "expertos internacionales".

El trabajo de recopilación, sistematización y análisis de la información de los testimonios recibidos por la CVR, así como su cruce con las demás bases de datos fue diseñado y realizado por un equipo de profesionales peruanos que trabajaron en el área de Sistemas de Información de la CVR. Como en todo proyecto importante y serio, es fundamental aprender de experiencias similares e incorporar sus enseñanzas para realizar un trabajo de la mejor calidad posible. Antes de la CVR existieron diversas Comisiones de la Verdad y proyectos parecidos en otros países del mundo (Chile, Argentina, Sud Africa, Guatemala, Kósovo, entre otros más). Varios de los profesionales que participaron en estos proyectos fueron invitados a compartir sus experiencias y conocimientos con la CVR Peruana, de tal manera que podamos incorporar este aprendizaje en nuestro propio trabajo. Patrick Ball y Jana Asher trabajaron en varios de estos proyectos (Patrick Ball en Haití, Sud Africa, El Salvador, Guatemala, Kosovo y Sierra Leone; Jana en Sierra Leone y Kosovo) y por ello fueron invitados a participar en el trabajo estadístico de la CVR junto con los miembros del equipo técnico de la CVR. Las conclusiones de este trabajo y sus implicancias fueron extensamente discutidas tanto por los comisionados, el equipo de profesionales de la CVR y otros académicos externos a la CVR antes de ser publicados en el informe final.

Habría que mencionar que en el 2002, Patrick Ball y Jana Asher recibieron el "Premio Especial por Logros Distinguidos" que otorga la Asociación Norteamericana de Estadística en virtud del trabajo que hicieron sobre la estimación de las muertes y desplazamientos durante el conflicto de Kósovo a fines de los años 90. Es ese trabajo se utilizaron muchas de las técnicas y métodos que fundamentan la estimación estadística elaborada por la CVR. En el 2002, Patrick Ball fue invitado por la fiscalía a presentar las conclusiones de esa investigación como parte de los informes periciales del juicio contra Slobodan Molisevic en el Tribunal Haya.

La experiencia de la CVR ha marcado nuevos hitos y pautas para el trabajo de Comisiones de la Verdad y proyectos similares en diversos países. De hecho muchos de los profesionales peruanos que trabajaron en la CVR se han convertido en "expertos internacionales" (varios ya lo eran antes de la CVR) compartiendo su experiencia en lugares tan diversos como Ghana, Timor Oriental, Sri Lanka, Colombia, Marruecos, Liberia y Sierra Leona. Lamentablemente mientras sigan habiendo conflictos en el mundo que provoquen masivas violaciones a los derechos humanos, seguirán siendo necesarios proyectos destinados a documentarlos y evitar que el mundo olvide las consecuencias de sus actos más terribles.

5.9) El investigador Daniel Manrique hizo una consultoría para la Presidencia del Consejo de Ministros el 2005 que demuestra que el número de muertos durante el Conflicto Armado Interno fue de sólo 13,218.

Esta "crítica" en realidad se trata de un rumor malintencionado cuyo origen se puede rastrear hasta una serie sistemática de declaraciones de un funcionario de la Presidencia del Consejo de Ministros en julio del 2006. La consultoría en cuestión realmente existió y fue realizada por uno de los autores de este documento (y del estudio estadístico de la CVR) pero la conclusión que se le atribuye es completamente falsa.

En realidad, el producto de esta consultoría fue una lista correspondiente a un *subconjunto* de los datos registrados por la CVR; aquellos que cumplían con un conjunto altamente restrictivo de características. La idea era que esta lista sirviera como insumo base para la creación de un registro nacional que pudiera ser usado para iniciar el proceso de reparaciones. La elaboración de esta lista no fue un

análisis estadístico (es más, ini si quiera contiene números!). El total al que se hace referencia (13,218) es un conteo simple de la cantidad de registros individuales en esa lista que, como ya se ha dicho, es incompleta por su propio diseño.

5.10) El experto peruano, Dr. Hugo Ñopo ha demostrado técnicamente / científicamente / contundentemente / etc. que las estimaciones de la CVR son incorrectas.

La mayor parte de los cuestionamientos técnicos (algunos razonables y otros más bien desinformados) contenidos y contestados en este documento han sido planteados por el Dr. Ñopo y, hasta donde tenemos conocimiento, estos son todos los que ha formulado. Como hemos intentado argumentar en este documento, la mayoría de estos cuestionamientos se basan en supuestos erróneos, y el resto son cuestiones interesantes que podrían ser exploradas en mayor profundidad; pero de allí a sostener que invalidan los resultados presentados hay un salto de proporciones cósmicas.

Como cualquier trabajo de naturaleza científica, la estimación de la CVR es debatible, perfectible e incluso refutable, pero a nuestro juicio ninguno de los argumentos esgrimidos hasta la actualidad ha tenido éxito demostrando o incluso sugiriendo su incorrección. Elevar críticas meramente retóricas sin sustento científico y técnico riguroso a la categoría de demostración nos parece francamente un exceso.

De hecho, la mayoría de las críticas "técnicas" al trabajo de la CVR han sido comentarios casuales, formulados en entrevistas o declaraciones a medios de comunicación, o publicadas en listas de discusión en Internet, que son citados una y otra vez. A pesar de que han pasado más de tres años desde la publicación del Informe Final y de que todos los datos utilizados para la estimación son de dominio público, no se han publicado artículos o documentos críticos al trabajo estadístico de la CVR que contengan argumentos técnicos *sustentados técnicamente* (no simples generalidades) en ninguna revista o medio científico serio o en ninguna otra parte.

El trabajo estadístico de la CVR, antes de ser publicado fue sometido a una revisión crítica por parte de diversos académicos expertos en temas estadísticos, entre ellos el propio presidente de la Asociación Norteamericana de Estadística. Además fue presentado como conferencia magistral en dos congresos peruanos de estudiantes de estadística (2003 y 2005) y más recientemente en el Congreso Latinoamericano de Probabilidad y Estadística Matemática realizado en Lima en 2007 (CLAPEM XIII) sin que se hayan formulado en esos foros cuestionamientos a la calidad del trabajo realizado.

Es cierto que un trabajo científico que usa complejas técnicas estadísticas es difícil de explicar para el público general. Todo lenguaje científico tiene su propia "jerga", sin embargo el que algunas personas no entiendan ese tipo de lenguaje, estén poco familiarizadas con él, o no estén dispuestas a tomarse el tiempo necesario para entenderlo no las autoriza a descartarlo como "mera decoración".

Bibliography

1. Comisión de la Verdad y Reconciliación (2003). *Informe Final, anexo 3 - ¿Cuántos Peruanos Murieron?: Estimación del Total de Víctimas Causadas por el Conflicto Armado Interno entre 1980 y el 2000*. Comisión de la Verdad y Reconciliación, Lima.

2. Efron, B., Tibshirani, R. J. (1993). *An Introduction to the Bootstrap*. Chapman & Hall, New York, NY.
3. Sanathanan, L (1972). *Estimating the Size of a Multinomial Population*. The Annals of Mathematical Statistics, Vol. 43, No. 1.
4. Fienberg, S.E. (1972) *The Multiple Recapture Census for Closed Populations and Incomplete 2^y Contingency Tables*. Biometrika Vol. 59, No. 3.
5. Bishop, Y., Fienberg, S.E. and Holland, P.H. (1975). *Discrete Multivariate Analysis. Theory and Practice*. MIT Press, Cambridge
6. Fienberg, S.E., Johnson, M.S., Junker, B.W. (1999) *Classical Multilevel and Bayesian Approaches to Population Size Estimation Using Multiple Lists-*. Journal of the Royal Statistical Society. Series A, Vol. 162, No 3.
7. Cormack, R.M (1992). *Interval estimates for mark-recapture studies of closed populations*. Biometrics, 48.
8. Centro de Estudios y Promoción del Desarrollo - DESCO (1989). *Violencia política en el Perú 1980-1988*. DESCO, Lima

Footnotes

[1](#). Es particularmente desafortunado que estas dos cifras coincidan además con la cifra de muertes que se manejaba antes del trabajo de la CVR. En realidad los tres números se refieren a cosas diferentes.