Ministerio de Cultura y Educación
Secretaría de Programación y Evaluación Educativa

Calidad de la Educación

Sistema Nacional de Evaluación
1er Operativo Nacional 1993

Diseño de la Muestra

Informe Técnico N°1

Lic. David Glejberman

I N D I C E

Página

1.OBJETIVOS

El objetivo principal de la Encuesta de Evaluación de la Calidad de la Educación 1993 es conocer, mediante técnicas de muestreo, el estado actual de los logros (rendimiento) obtenidos en las áreas instrumentales (Lengua y Matemática) por parte de los alumnos de la enseñanza primaria y media al finalizar cada nivel, de acuerdo con las determinantes de la Ley Federal de Educación del 24/3/93 (Arts. 48,49 y 50).

Se consideran objetivos complementarios aportar al conocimiento de la asociación (correlación) del rendimiento escolar de los alumnos con variables relacionadas con la conformación de sus hogares, la educación de los padres, los antecedentes escolares de los alumnos, las características de los docentes del curso y del director del establecimiento.

La utilización del muestreo y no de la enumeración completa (censo) está determinada por los objetivos que se persiguen, el presupuesto disponible y la oportunidad deseada en la presentación de los resultados. En virtud de los objetivos recién enunciados, en opinión del Consultor, el muestreo es el procedimiento más eficiente, tanto en términos de costo como de oportunidad.

2.UNIVERSO A INVESTIGAR

El universo a investigar es el conjunto de alumnos que en la Argentina cursan, en 1993, el último grado de la enseñanza primaria común y de la enseñanza media en las modalidades Bachillerato, Comercial y Técnica (Industrial y Agraria).

Se excluyen del universo la enseñanza primaria especial y la de adultos; y en el caso de la enseñanza media se excluyen ciertas modalidades (especial, de adultos, artística, etc.).

Una vez definido el procedimiento para el levantamiento de la información (toda la muestra se ejecuta en un mismo día para cada prueba, en dos días consecutivos), se encuentra que el muestreo por conglomerados (utilizando la sección escolar o la división como conglomerado) resulta más beneficioso que otros posibles diseños (ver justificación en la sección 6.), excepto por la gran variabiidad del tamaño de las secciones escolares en el ámbito rural. En la mayor parte de las jurisdicciones con establecimientos rurales se encuentra un número importante de secciones con muy pocos alumnos en el 7o. grado (generalmente establecimientos con personal único), respecto de los cuales la relación costo-cantidad de información, en el caso de pertenecer a la muestra, se vuelve ineficiente. Razones de costo y relacionadas con la cantidad de recursos humanos disponibles para el levantamiento

de la información, determinan la conveniencia de excluir del universo las secciones con menos de 5 alumnos. En tal caso, en las jurisdicciones con una matrícula escolar relevante en el ámbito rural, se estima que alrededor del 30% de las secciones quedan excluídas de la investigación, pero que en términos del número de alumnos representa en el orden del 5% de la matrícula de cada jurisdicción.

En virtud de las razones expuestas, el equipo técnico de la Dirección Nacional de Evaluación decidió, con el acuerdo de las jurisdicciones, redefinir el universo a investigar excluyendo los alumnos que en los dos niveles cursan en 1993 en secciones o divisiones con 4 o menos estudiantes.

En el caso de la enseñanza media, la exclusión de las divisiones con hasta 4 alumnos prácticamente no modifica la definición del universo porque la existencia de tales divisiones es excepcional.

En el siguiente cuadro se presenta una estimación del tamaño del universo a investigar a partir de los datos disponibles a la fecha de este informe.

Cuadro 1: Número de alumnos matriculados en 1993 en el último grado en secciones o divisiones de 5 o más estudiantes, por nivel de enseñanza según jurisdicción.

JURISDICCION	NIVEL DE ENSEÑANZA PRIMARIA	MEDIO
TOTAL	**597.323**	**249.651**
1. Capital Federal	35.552*	31.076*
2. Gran Buenos Aires	139.876	58.647*
3. Buenos Aires (demás partidos)	91.233	31.671*
4. Catamarca	5.669	2.191
5. Córdoba	57.373	23.524*
6. Corrientes	16.035	5.099
7. Chaco	13.949	4.562
8. Chubut	7.498	2.720
9. Entre Ríos	20.725	8.412
10. Formosa	6.524*	2.727*
11. Jujuy	12.827	6.067
12. La Pampa	4.826	2.155
13. La Rioja	4.172	1.849
14. Mendoza	26.051	10.081
15. Misiones	15.616	5.074
16. Neuquén	8.471	2.560
17. Río Negro	7.645*	3.692*
18. Salta	18.243	6.866
19. San Juan	10.568	4.801*
20. San Luis	5.329	1.855
21. Santa Cruz	2.683*	1.330*
22. Santa Fe	47.577	19.229*
23. Santiago del Estero	13.475	5.156
24. Tierra del Fuego	1.206	298
25. Tucumán	24.200	8.009

Nota: * Estimaciones a partir del Censo de Población 1991 (INDEC).

3. UNIDADES DE OBSERVACION

De acuerdo con los objetivos ya planteados, las unidades a investigar son:

- el alumno
- la familia del alumno
- los maestros o profesores del alumno
- el director del establecimiento donde cursa el alumno

Para obtener la información relativa a cada una de estas unidades de observación, se han confeccionado los siguientes instrumentos:

NIVEL	UNIDAD DE OBSERVA-CION	CUESTIONARIOS	CONCEPTO
PRIMARIO	ALUMNOS	PPm	Prueba Primario matemática
PRIMARIO	ALUMNO	PPl	Prueba Primario de lengua
PRIMARIO	ALUMNO	CAP-U,CAP-R	Cuestionario Alumno Primario
PRIMARIO	FAMILIA	CF-U, CF-R	Cuestionario Familia
PRIMARIO	MAESTRO	CMm-U	Cuestionario Maestro matemática
PRIMARIO	MAESTRO	CMl-U	Cuestionario Maestro lengua
PRIMARIO	MAESTRO	CM-R	Cuestionario Maestro
PRIMARIO	DIRECTOR	CDP-U,CDP-R	Cuestionario Director Primario
MEDIO	ALUMNO	PMm	Prueba Medio matemática
MEDIO	ALUMNO	PMl	Prueba Medio lengua
MEDIO	ALUMNO	CAM	Cuestionario Alumno Medio
MEDIO	PROFESOR	CPm	CuestionarioProfesor matemática
MEDIO	PROFESOR	CPl	Cuestionario Profesor lengua
MEDIO	DIRECTOR	CDM	Cuestionario Director Medio

4. NIVELES DE DESAGREGACION DE LOS RESULTADOS

De acuerdo con las estimaciones realizadas a la fecha, la muestra diseñada podría proporcionar información sobre las cantidades aproximadas que se muestran en el siguiente cuadro.

Cuadro 2: Cantidad aproximada de unidades de observación en la muestra por nivel de enseñanza según tipo de unidad

UNIDADES DE OBSERVACION	NIVEL DE ENSEÑANZA	
	PRIMARIO	MEDIO
ALUMNOS	13.500	10.000
FAMILIAS	13.500	10.000
DOCENTES	1.350	1.000
DIRECTORES	900	500

Estos tamaños resultan, de acuerdo con el diseño que se comenta en la sección 6, del objetivo principal de la investigación (conocer los logros en las áreas instrumentales) en relación con la unidad de observación "alumno" para una cierta precisión y seguridad deseadas, y de las características propias del diseño muestral (estratificado y por conglomerados). Resultan así estimaciones confiables para los parámetros investigados en diferentes niveles de desagregación según la unidad de observación de que se trate. El siguiente cuadro resume las diferentes situaciones que pueden plantearse para cada unidad de observación.

Cuadro 3: Niveles de desagregación por unidad de observación

NIVEL DE ENSEÑANZA	UNIDAD DE OBSERVACION	NIVELES DE DESAGREGACION
PRIMARIO	ALUMNOS	NIVEL NACIONAL
PRIMARIO	ALUMNOS	JURISDICCION
PRIMARIO	ALUMNOS	AMBITO
PRIMARIO	ALUMNOS DE JURISDIC. CON MUESTRA EXTENDIDA	JURISD. Y AMBITO
PRIMARIO	FAMILIAS	NIVEL NACIONAL
PRIMARIO	FAMILIAS	JURISDICCION
PRIMARIO	FAMILIAS	AMBITO
PRIMARIO	FAMILIAS DE JURISDIC. CON MUESTRA EXTENDIDA	JURISD. Y AMBITO
PRIMARIO	MAESTROS MATEMATICA	NIVEL NACIONAL URBANO
PRIMARIO	MAESTROS LENGUA	NIVEL NACIONAL URBANO
PRIMARIO	MAESTROS RURALES	NIVEL NACIONAL RURAL
PRIMARIO	DIRECTORES	NIVEL NACIONAL
PRIMARIO	DIRECTORES	AMBITO
MEDIO	ALUMNOS	NIVEL NACIONAL
MEDIO	ALUMNOS	JURISDICCION
MEDIO	ALUMNOS	MODALIDAD
MEDIO	PROFESORES MATEMATICA	NIVEL NACIONAL
MEDIO	PROFESORES LENGUA	NIVEL NACIONAL
MEDIO	DIRECTORES	NIVEL NACIONAL

Cualquier forma de desagregación más fina que las aquí planteadas generará resultados poco confiables en virtud del reducido tamaño de la muestra. No obstante, de acuerdo con el plan de estimadores propuesto en la sección 8, será posible calcular la pérdida de precisión que se derivaría de un mayor nivel de desagregación para la respectiva unidad de observación y variables a investigar (ver en particular los comentarios sobre la construccción de intervalos de confianza en la sección 9).

5. MARCO MUESTRAL

De acuerdo con los objetivos planteados se hace necesario disponer de un marco muestral a los siguientes efectos:

- determinar el tamaño de la muestra
- sortear la muestra de acuerdo con las características del diseño propuesto
- ponderar los resultados que se obtengan del relevamiento, si la muestra no es autoponderada.

En virtud de las ventajas que se derivan de la estratificación a priori, el marco muestral debería ser también estratifica-

do, en lo posible con los mismos criterios planteados para la desagregación de los resultados: nivel geográfico, ámbito (en el nivel primario) y modalidad (en el nivel medio).

Ante la inexistencia de un marco previo, se tomó la decisión de elaborarlo con la colaboración de las autoridades de la Educación de las respectivas jurisdicciones. A estos efectos se cursó el correspondiente pedido - vía la Red Federal de Educación - para que las jurisdicciones proporcionaran los datos de acuerdo con el siguiente esquema:

PRIMARIO

DEPARTAMENTO	AMBITO	NOMBRE DEL ESTABLECIM.	DIRECCION	CANTIDAD AL 31/3/93)	
				Secc. 7o.	Alumnos 7o.

MEDIO

DEPARTAMENTO	MODALI	NOMBRE DEL ABLECIM.	DIRECCION	CANTIDAD A 31/3/93	
				último grado	último grad

Habría sido deseable contar con mayor información en el marco, relativo al rendimiento escolar y su variabilidad, o bien alguna otra variable con alta correlación con aquélla. Todo intento en este sentido habría sido infructuoso en virtud del tiempo de que se disponía para la tarea y, en muchos casos, de la inexistencia de información de este tipo. Obsérvese que estos datos, al nivel muestral, estarán disponibles para futuras investigaciones, una vez procesado el operativo de 1993. Como consecuencia de ello, y con información sobre los costos (y las diferencias de costo por estrato) será posible implementar un diseño muestral basado en el muestreo estratificado óptimo (con afijación de Neyman).

Los marcos muestrales provenientes de las jurisdicciones fueron evaluados comparándolos con los datos sobre educación provenientes del Censo de Población y Vivienda de 1991 (fuente: INDEC, sobre cuadros aún no publicados a la fecha de este informe), y re-elaborados para estos fines.

Finalmente se procedió a depurar los marcos eliminando los grupos con menos de 5 alumnos, y de ello resultan los datos que se presentaron en el Cuadro 1. Los datos que figuran con asterisco en dicho cuadro responden a tres situaciones bien diferenciadas a la fecha de este informe:

- los marcos depurados proporcionaron datos agregados inferiores a los totales que provienen del Censo de Población de 1991
- el trabajo de depuración estaba pendiente

- las jurisdicciones no habían enviado aún la información pertinente

En el primer caso se solicita a las respectivas jurisdicciones que revisen y completen la información que proporcionaron del marco muestral, pues de no corregirlo se generarían "errores de cobertura" (ciertas secciones o divisiones tendrían probabilidad nula de ser elegidas por su no inclusión en el marco) y los datos agregados no serían válidos para ponderar los resultados de la jurisdicción a los efectos de obtener las estimaciones al nivel nacional.

En el segundo caso el trabajo de depuración estaba pendiente por demoras en el envío de los datos o porque la presentación de los mismos hizo más dificultosa la tarea de depuración.

El tercer caso es el más complicado, porque no tiene una solución aceptable desde el punto de vista del muestreo. Si no se dispone del marco muestral de una jurisdicción, entonces no es posible seleccionar una muestra probabilística de la misma y, en consecuencia, los resultados que se obtengan serán representativos de un universo que se define con exclusión de aquella jurisdicción.

Esta situación, obviamente no deseada, está contemplada en el diseño muestral que se propone en la sección siguiente.

6. DISEÑO MUESTRAL PROPUESTO

El diseño muestral que se propone es, en general, el muestreo aleatorio estratificado* en una sola etapa de selección. Esto es válido para las siguientes unidades de observación: docentes (maestros y profesores) y directores en ambos niveles de enseñanza. En cambio, el diseño es "complejo" en el caso de los alumnos (ambos niveles) y las familias (nivel primario). Se denomina "complejo" a un diseño muestral que combina dos o más técnicas simples de muestreo. En el caso de las unidades de observación alumnos y familias el diseño muestral propuesto combina la estratificación previa con el muestreo por conglomerados.

Dado que el objetivo de la investigación refiere principalmente a los alumnos, el diseño muestral estratificado y por conglomerados se propuso con el objeto de realizar inferencia acerca de los dos universos de alumnos: del último

---

* El muestreo es estratificado cuando previo a la selección de la muestra el universo es dividido en clases separadas (estratos), y la muestra se elige independientemente en cada estrato por el procedimiento conocido como muestreo aleatorio simple sin resposición.

grado del nivel primario y del último grado del nivel medio. En el caso de las restantes unidades de observación - familias, maestros, profesores y directores - el diseño muestral es consecuencia del que se define para las poblaciones de alumnos, tal como se explica al final de esta sección.

6.1 DISEÑO MUESTRAL PARA LOS ALUMNOS DEL NIVEL PRIMARIO

Este diseño es estratificado y por conglomerados. La estratificación es doble: por jurisdicción y por ámbito. El conglomerado es la sección escolar. Seleccionada una sección, todos sus alumnos quedan automáticamente incluídos en la muestra, esto es, se realiza un censo del conglomerado (una sola etapa de selección). La elección del diseño (de conglomerados y utilizando como unidad primaria la sección escolar) no es arbitraria sino que responde adecuadamente a las características del operativo de evaluación de la calidad educativa. Podría haberse utilizado como unidad primaria el establecimiento escolar en lugar de la sección escolar, pero ésta presenta diversas ventajas tanto del punto de vista del muestreo como del propio trabajo de campo. Basta señalar que las secciones tienen tamaños más similares entre sí (dentro de cada estrato) que las cantidades de alumnos del 7º grado en cada establecimiento, y que la administración de la pruebas y cuestionarios de una sección escolar es una tarea que puede desarrollar razonablemente una sola persona en el horario de clases.

La opción por la enumeración completa dentro del conglomerado, por oposición a la subnumeración (muestreo en dos etapas) se justifica otra vez tanto por el lado del muestreo como por el trabajo de campo. Desde el punto de vista del muestreo, al interior de la sección escolar es de esperar una razonable variabilidad del rendimiento escolar antes que uniformidad (alumnos muy buenos, buenos, regulares y malos), en consecuencia, el censo de la sección proporciona información más rica que una muestra de la misma. Y esta mayor riqueza de información no genera mayores costos en el trabajo del encuestador que debe administrar pruebas y cuestionarios (obviamente sí se tienen mayores costos de impresión de cuestionarios y de procesamiento).

La estratificación de los alumnos es doble: por jurisdicción y por ámbito. En estrictez, y para contemplar las desagregaciones requeridas, así como los problemas que podrían derivarse de la no inclusión de alguna provincia en la investigación, las jurisdicciones funcionan como subuniversos, para cada uno de los cuales se tienen dos

estratos: urbano y rural (con la sola excepción de Capital Federal, para la que no existe el estrato rural).

La determinación del tamaño de la muestra se realiza independientemente en cada jurisdicción a los efectos de obtener 5 puntos de precisión con una seguridad del 95% en la estimación del rendimiento escolar esperado en las pruebas (se supone un puntaje entre 0 y 100).

En el caso de las jurisdicciones con muestra "extendida", los estratos, urbano o rural, son tratados como universos independientes. En estos casos, la precisión planeada es de 10 puntos para el ámbito rural y de 5 puntos en el ámbito urbano.

En el caso de las jurisdicciones con muestra no extendida, la determinación del tamaño de la muestra se realiza mediante el muestreo aleatorio estratificado proporcional (en proporción al número de alumnos).

Como puede observarse, la consideración de las jurisdicciones como subuniversos, y la dualidad de criterios para la determinación de la muestra en cada estrato (muestra extendida - muestra no extendida) no resulta en una muestra autoponderada por cuanto no se respeta la proporcionalidad al nivel nacional. Ello implica que el cálculo de los estimadores debe realizarse separadamente por estrato y luego ponderarse adecuadamente para obtener estimaciones de los parámetros de subuniversos y del universo total.

6.2 DISEÑO MUESTRAL PARA LOS ALUMNOS DEL NIVEL MEDIO

Este diseño es semejante al de los alumnos del nivel primario. Es estratificado y por conglomerados. El conglomerado es la división, y valen para ella los mismos comentarios realizados con relación a la sección escolar. La estratificación es doble: por jurisdicción y por modalidad. Se investigan tres modalidades: bachillerato, comercial y técnica (industrial y agraria).

La determinación del tamaño de la muestra se realiza independientemente en cada jurisdicción a los efectos de obtener 5 puntos de precisión con una seguridad del 95% en la estimación del rendimiento escolar promedio esperado en las pruebas (también se supone un puntaje entre 0 y 100).

La asignación del tamaño de la muestra en cada estrato (modalidad) se realiza mediante el muestreo aleatorio estratificado proporcional (en proporción al número de alumnos).

También en este caso, al considerarse las jurisdicciones como subuniversos, la muestra no resulta autoponderada, lo que exige un cuidado especial en el cálculo de los estimadores y en la identificación de cuestionarios y pruebas para asignarlos al estrato correspondiente.

6.3 DISEÑO MUESTRAL PARA FAMILIAS, MAESTROS, PROFESORES Y DIRECTORES.

Para estas unidades de observación, los respectivos diseños muestrales quedan determinados por los correspondientes diseños de los universos principales (alumnos del nivel primario y alumnos del nivel medio).

En lo que se refiere a las familias, habrá tantas unidades en la muestra como alumnos, además, con el mismo procedimiento de selección. Puede ocurrir, con una probabilidad despreciable, que una misma familia aparezca dos veces en la muestra (porque tiene dos hijos en el mismo grado que salieron sorteados en la muestra), en cuyo caso se tendrá una muestra de familias con elementos repetidos. Si se diera el caso, se solicitará a la familia que responda dos veces al respectivo cuestionario, por cuanto lo que interesa es relacionar las variables de la familia con las del alumno. A los efectos del muestreo, la muestra de familias será tratada, aún en este caso, como si fuera sin resposición.

En consecuencia, el diseño muestral para las familias es estratificado y por conglomerados, tal como en el caso de los alumnos del nivel primario.

En cuanto a los maestros, en los tres casos (de matemática, de lengua y rural) se tiene uno por cada sección escolar en la muestra, y como las secciones se eligen por muestreo aleatorio estratificado, la muestra de maestros (en cada especialidad) es una muestra aleatoria estratificada de maestros.

Si interesa investigar la categoría "maestro urbano" que incluye a los maestros de matemática y de lengua, habrá que tener en cuenta que los mismos no son seleccionados independientemente, sino por pares, por estar a cargo de la misma sección escolar. En consecuencia, para esta categoría el diseño muestral es estratificado y por conglomerados (con el tamaño del conglomerado igual a dos, en todos los casos).

Con relación a los profesores del nivel medio, valen los mismos comentarios que para los maestros. El diseño muestral para los profesores de matemática (así como para el profesor de lengua) es el muestreo aleatorio estratificado. Si interesa investigar la categoría "profesor" (independientemente de la especialidad) el diseño es estratificado y por conglomerados (con tamaño igual a dos).

En el caso poco probable que un profesor salga sorteado dos veces en la muestra (porque es profesor a la vez de dos divisiones seleccionadas en la muestra) deberá responder dos veces al cuestionario.

En virtud del procedimiento de selección de los conglomerados (aleatorio simple sin reposición en cada estrato) es muy poco probable que un director salga sorteado dos veces en la muestra. Se ocurriera este hecho, debería repetir el cuestionario respectivo por cada sección (división) seleccionada en la muestra perteneciente al establecimiento que dirige. De acuerdo con lo ya visto para maestros y profesores, el diseño muestral para los directores es el muestreo aleatorio estratificado.

7. AFIJACION DE LA MUESTRA

Se entiende por afijación de la muestra el problema de determinar el tamaño de la muestra en cada estrato del universo. La muestra se determina a partir de los universos de alumnos, para obtener cierta precisión y seguridad con relación al rendimiento escolar promedio. Excepto en el caso de

jurisdicciones con muestra extendida, el procedimiento para la afijación de la muestra es el mismo para los dos niveles de enseñanza. A los efectos de la determinación del tamaño de la muestra se supone que para la estimación del rendimiento escolar promedio esperado se utilizan los clásicos estimadores lineales insesgados (para un mayor detalle de las opciones y el levantamiento de este supuesto, ver sección 8).

7.1 AFIJACION DE LA MUESTRA NO EXTENDIDA

Notación: X = rendimiento escolar

μ = media poblacional (parámetro a estimar)

N = tamaño poblacional (Nº de alumnos del universo)

N_h = tamaño poblacional del estrato h

M_h = Nº de conglomerados en el estrato h

$T_h = N_h/M_h$ = tamaño medio de los conglomerados en el estrato h

σ_h^2 = varianza poblacional de X en el estrato h

ρ_h = coeficiente de correlación intraconglomerado en el estrato h.

m_h = tamaño de la muestra de conglomerados en el estrato h

n_h = tamaño de la muestra de alumnos en el estrato h

$\tilde{\mu}_h$ = estimador de la media poblacional en el estrato h

K = número de estratos en la población

T_{jh} = tamaño del conglomerado j en el estrato h

Planteo del problema: determinar el tamaño de la muestra en cada estrato para estimar μ con una precisión ϵ y

[illegible] $P(|\tilde{\mu} - \mu| < \epsilon) \geq 0.95$

Se define: $\tilde{\mu} = \sum_{h=1}^{k} \frac{N_h}{N} \tilde{\mu}_h$, *donde* $\tilde{\mu}_h = \frac{1}{m_h} \sum_{j=1}^{m_h} \bar{X}_{jh}$

donde $\bar{X}_{jh} = \frac{\sum_{i}^{T_{jh}} X_{ijh}}{T_{jh}}$

es el promedio de la variable X en el conglomerado j del estrato h de la muestra.

En todos los casos se supone que $\tilde{\mu}_h \rightarrow N(\mu_h, V(\tilde{\mu}_h))$, donde:

$V(\tilde{\mu}_h) = \frac{\sigma_h^2}{n_h} [1 + (T_h - 1)\, \rho_h] \frac{N_h - n_h}{N_h - 1}$, y como consecuencia de ello:

$$V(\tilde{\mu}) = \sum_{1}^{k} (\frac{N_h}{N})^2 \frac{\sigma_h^2}{n_h} [1 + (T_h - 1)\, \mu_h] \frac{N_h - n_h}{N_h - 1}$$

Ahora, si se parte del planteo $P(|\tilde{\mu} - \mu| < \epsilon) = 0.95$, , entonces

$$P\left(\tilde{\mu}^* < \frac{\epsilon}{\sqrt{V(\tilde{\mu})}} \right) = 0.975$$

$$\frac{\epsilon}{\sqrt{V(\tilde{\mu})}} = 1.96 \rightarrow \boxed{V(\tilde{\mu}) = \left(\frac{\epsilon}{1.96}\right)^2} \quad (1)$$

Si se desea una muestra de conglomerados proporcional al número de alumnos en cada estrato, entonces:

$$\boxed{m_h = m\frac{N_h}{N}} \quad (2)$$

donde m = total de conglomerados en la muestra = $\sum_{1}^{k} m_h$.

$V(\tilde{\mu})$ puede descomponerse como la suma de productos de dos factores a saber:

$$V(\tilde{\mu}) = \sum_{1}^{K} \left\{ \left(\frac{N_h}{N} \right)^2 \sigma_h^2 \left[1 + (T_h - 1)\rho_h \right] \right\} \cdot \left\{ \frac{1}{n_h} \frac{N_h - n_h}{N_h - 1} \right\}$$

Con relación al primer factor, a falta de otra información, se supone que $\sigma_h^2 = \sigma^2$ $\forall h$. En el caso que se dispusiera de información sobre las varianzas por estrato, entonces se podría realizar una afijación más eficiente (óptima o de Neyman).

También se supondrá que $[1 + (T_h - 1)\ \rho_h] = 3.5$ $\forall h$ ("efecto de diseño" de KISH), lo que implica que la correlación intraclase del rendimiento escolar es del orden de 0.10 para grupos de alrededor de 25 alumnos (primario-urbano y medio en las tres modalidades y $\rho_h \cong 0.5$ para secciones con $T_h \cong 6$ (primario - rural).

Con relación al segundo factor, y teniendo en cuenta que $n_h \cong m_h\ T_h$ resulta:

$$\frac{1}{n_h} \frac{N_h - n_h}{N_h - 1} \cong \frac{1}{m T_h \frac{N_h}{N}} \frac{N_h - m T_h \frac{N_h}{N}}{N_h} = \frac{N}{m T_h N_h} - \frac{1}{N_h}$$

Entonces: $$V(\tilde{\mu}) = \sum_{1}^{k} \left(\frac{N_h}{N} \right)^2 \sigma^2 \times 3.5 \left\{ \frac{N}{m T_h N_h} - \frac{1}{N_h} \right\} = \frac{3.5\sigma^2}{N} \left[\sum_{1}^{k} \frac{1}{m} \frac{N_h}{T_h} - \sum_{1}^{k} \frac{N_h}{N} \right]$$

$$= \frac{3.5\sigma^2}{N} \left[\sum_{1}^{k} \frac{M_h}{m} - 1 \right] \rightarrow V(\tilde{\mu}) = \frac{3.5\sigma^2}{N} \left[\frac{M-1}{m} \right] = \left(\frac{\epsilon}{1.96} \right)^2 \text{ donde } M = \sum_{1}^{k} M_h.$$

Resulta: $$\frac{M}{m}-1=\left(\frac{\epsilon}{1.96}\right)^2\frac{N}{3.5\sigma^2}\rightarrow \boxed{m=\frac{M}{1+\left(\frac{\epsilon}{1.96}\right)^2\frac{N}{3.5\sigma^2}}}$$

Determinada m por la fórmula obtenida, la afijación de la muestra por estratos se realiza utilizando (2).

Con relación a los resultados recién obtenidos, es necesario tener en cuenta las siguientes observaciones:

i) la fórmula para la determinación del tamaño de la muestra se ha simplificado notablemente como consecuencia del supuesto $\sigma_h^2=\sigma^2$. Cuando sea posible levantar este supuesto (porque se cuenta con la información pertinente) será necesario volver a calcular la fórmula que determina el tamaño de la muestra. Cuando ello sea posible se sugiere cambiar el criterio de asignación proporcional por el de asignación óptima.

ii) Para estimar σ^2 en cada nivel de enseñanza se utilizaron los resultados obtenidos en el operativo de evaluación de la calidad de la educación realizado en 1991. Cambiando convenientemente la escala de puntajes (para llevarlo a un máximo de 100) y promediando los resultados de lengua y matemática, se obtuvo:

Nivel	$\tilde{\sigma}^2$
PRIMARIO	750
MEDIO	800

iii) En la determinación de los tamaños de muestra en 1993, en casi todos los casos, se trabaja con ϵ = 5 en el supuesto de que las pruebas de evaluación puntuarán en una escala de 1 a 100. En realidad se desea una precisión del 5% al puntaje más alto de la prueba.

iv) En la muestra de 1993 se trabaja con ϵ = 5 a nivel de las jurisdicciones, y se espera que ello conduzca, a nivel nacional a un ϵ = 2. Para verificarlo, a

posteriori de haber realizado el operativo, se estimará:

$\epsilon = 1.96\sqrt{V(\tilde{\mu})}$ a través de $\tilde{\epsilon} = 1.96\sqrt{\tilde{V}(\tilde{\mu})}$

para lo cual será necesario estimar σ_h^2 en cada estrato.

$$\tilde{V}(\tilde{\mu}) = \sum_{1}^{k} \left(\frac{N_h}{N}\right)^2 \frac{\tilde{\sigma}_h^2}{n_h} \left[1 + (\tilde{T}_h - 1)\,\tilde{\rho}_h\right] \frac{N_h - n_h}{N_h - 1}$$

donde: T_h se estimará a partir del tamaño medio de los conglomerados en la muestra, mientras que σ_h^2 y ρ_h se podrán estimar a partir del procedimiento simplificado que se propone en la sección 9.

v) Dado que el marco muestral es al 31/3/93 y el operativo se realiza en el mes 11/93, una parte de la matrícula al 31/3 no estará presente en las pruebas, por deserción o por ausencia temporaria. Para contemplar esta omisión en la determinación del tamaño de la muestra se supone en todos los estratos que el tamaño efectivo de los conglomerados en la muestra será un 15% inferior que en el marco. Ello equivale a incrementar el tamaño de la muestra aproximadamente un 18% en términos del número de conglomerados para asegurarse un tamaño adecuado de la muestra de alumnos.

7.2 AFIJACION DE LA MUESTRA EXTENDIDA

En las jurisdicciones que acordaron con la Dirección Nacional de Evaluación una muestra ampliada para el nivel primario, el objetivo consiste en obtener resultados confiables para la jurisdicción en cada ámbito. Ahora los estratos funcionan como subuniversos. Utilizando la misma notación que en 7.1 (sin el subíndice h que representa al estrato) el tamaño de la muestra en cada jurisdicción y ámbito se obtiene de acuerdo con el siguiente procedimiento.

Como en el caso anterior,

$$P\left(\left| \tilde{\mu} - \mu \right| < \epsilon \right) = 0.95$$

conduce a: $$\boxed{V(\tilde{\mu}) = \left(\frac{\epsilon}{1.96}\right)^2} \quad (1)$$

Al tratarse de una muestra aleatoria simple sin reposición de conglomerados en el subuniverso, resulta:

$$V(\tilde{\mu}) = \frac{\sigma^2}{n}\left[1 + (T-1)\rho\right]\frac{N-n}{N-1}$$

Si suponemos $n \cong mT \Rightarrow V(\tilde{\mu}) = \frac{\sigma^2}{m.T} 3.5 \frac{N-mT}{N-1}$

Si se admite N - 1 ≅ N, entonces:

$$V(\tilde{\mu}) \cong 3.5\,\sigma^2\left[\frac{1}{m.T} - \frac{1}{N}\right] = \left(\frac{\epsilon}{1.96}\right)^2$$

De donde $$\boxed{m = \frac{M}{1 + \left(\frac{\epsilon}{1.96}\right)^2 \frac{N}{3.5\,\sigma^2}}}$$

Obsérvese que la fórmula es la misma que la obtenida en 7.1, con la única diferencia en la interpretación de m, M y N. En 7.1, m es el número de conglomerados en la muestra de la jurisdicción mientras que en 7.2 m es el número de conglomerados en la muestra del ámbito respectivo dentro de la jurisdicción. Otro tanto ocurre con la interpretación de M y N en cada caso.

También en este caso, en la determinación del tamaño de la muestra de conglomerados se corrige suponiendo una omisión del 15% de los alumnos respecto de la matrícula inicial del marco muestral.

8. PLAN DE ESTIMADORES Y SUS VARIANZAS

8.1 ALUMNOS Y FAMILIAS

Con relación al universo de alumnos en cada nivel, suponemos que interesa estimar:

$$\mu = \frac{\sum_{1}^{N} X_i}{N} =$$

media poblacional de la variable X, la que también puede obtenerse a través de los estratos y conglomerados así:

$$\mu = \frac{\sum_{h}^{k} \sum_{j}^{M_h} \sum_{i}^{T_{jh}} X_{ijh}}{N}$$

Si lo que interesa conocer es la proporción (p) de alumnos que en la población poseen cierto atributo, entonces basta asignar a la variable X_i el valor 1 si el individuo i posee el atributo, y 0 si no lo posee. En tal caso, el parámetro p se define de la misma forma que μ.

Para proponer los estimadores adecuados para μ (y p) se debe tener en cuenta que los conglomerados en la población (y también en la muestra) son de tamaños desiguales. Esta variabilidad en el tamaño de los conglomerados (sobre todo en las zonas rurales) hace más eficiente el uso de los estimadores de razón en lugar de los clásicos estimadores lineales (o de "expansión simple"). De esta forma, se propone la utilización de estimadores sesgados, con un sesgo despreciable en muestras grandes, porque presentan un menor error cuadrático medio que los estimadores lineales insesgados.

En la notación que sigue el subíndice h representa al estrato, j al conglomerado e i al alumno. En cada caso, para cada nivel de desagregación, h representa los estratos que definen dicho nivel. Así, si la media a estimar es la de todos los alumnos del primario, entonces los estratos que definen dicho nivel. Así, si la media a estimar es la de todos los alumnos del primario, entonces los estratos a

incluirse son los de todas las jurisdicciones para ambos ámbitos. Si la media a estimar es la de todos los alumnos del bachillerato, entonces los estratos a incluirse son los que corresponden a dicha modalidad en todas las jurisdicciones. Si interesa la media de una sola jurisdicción (con muestra extendida) de los alumnos del primario en el ámbito urbano, entonces en las fórmulas que se proponen a continuación se incluirá un solo estrato (el que corresponde a esa jurisdicción en el ámbito urbano).

Con estas precisiones, el estimador que se propone para la media es:

$$\tilde{\mu} = \sum_{h=1}^{k} \frac{N_h}{N} \tilde{\mu}_h, \text{ con } \tilde{\mu}_h = \frac{\sum_{j}^{m_h} \sum_{i}^{T_{jh}} X_{ijh}}{\sum_{j}^{m_h} T_{jh}}$$

donde: X_{ijh} es el valor de la variable X del individuo i en el conglomerado j de la muestra en el estrato h.

T_{jh} es el tamaño del conglomerado j de la muestra en el estrato h.

De acuerdo con los resultados del muestreo aleatorio estratificado, la varianza de $\tilde{\mu}$ es:

$$V(\tilde{\mu}) = \sum_{h=1}^{k} \left(\frac{N_h}{N}\right)^2 V(\tilde{\mu}_h)$$

donde cada $V(\tilde{\mu}_h)$ se obtiene tomando en consideración que $\tilde{\mu}_h$ es un estimador de razón a partir de un muestreo aleatorio simple sin reposición de conglomerados. La fórmula de $V(\tilde{\mu}_h)$ es harto complicada, pero se tiene una buena aproximación a partir de la siguiente fórmula (W. Cochran: "Técnicas de Muestreo"):

$$V(\tilde{\mu}_h) \cong \left(1 - \frac{m_h}{M_h}\right) \frac{1}{m_h \, T_h^2} \frac{\sum_j^{M_h} T_{jh}^2 \, (\bar{X}_{jh} - \mu_h)^2}{M_h - 1}$$

donde: $\bar{X}_{jh} = \dfrac{\sum_i^{T_{jh}} X_{ijh}}{T_{jh}} =$ promedio de la variable X en el conglomerado j para la población del estrato h

$T_h = \dfrac{\sum_j^{M_h} T_{jh}}{M_h} =$ tamaño medio de los conglomerados del estrato h para toda la población

De estos datos, obviamente, solo se conoce T_h para la matrícula inicial al 31/3/93, o bien la versión corregida por omisión (-15%). Por tanto, para aproximarse a $V(\tilde{\mu}_h)$ será necesario estimar T_h y el último factor. Se propone como estimador de $V(\tilde{\mu}_h)$ la expresión siguiente:

$$\tilde{V}(\tilde{\mu}_h) = \left(1 - \frac{m_h}{M_h}\right) \frac{1}{m_h \tilde{T}_h^2} \frac{\sum_j^{m_h} T_{jh}^2 (\bar{x}_{jh} - \tilde{\mu}_h)^2}{m_h - 1}$$

donde: $\bar{x}_{jh} = \dfrac{\sum_i^{T_{jh}} x_{ijh}}{T_{jh}} =$ promedio de la variable X en el conglomerado j para la muestra del estrato h

$$\tilde{T}_h = \frac{\sum_j^{m_h} T_{jh}}{m_h} =$$ tamaño medio de los conglomerados del estrato h en la muestra

De lo anterior se deduce que un estimador apropiado de la varianza de $\tilde{\mu}$ (media ponderada de varios estratos) es:

$$\tilde{V}(\tilde{\mu}) = \sum_{h=1}^{K} \left(\frac{N_h}{N}\right)^2 \left(1 - \frac{m_h}{M_h}\right) \frac{1}{m_h \tilde{T}_h^2} \frac{\sum_j^{m_h} T_{jh}^2 (\bar{x}_{jh} - \tilde{\mu}_h)^2}{m_h - 1}$$

En el Anexo 1 de este documento se presenta otro posible estimador para μ (el de "expansión simple") con su respectiva varianza y correspondiente estimador.

8.2 MAESTROS, PROFESORES Y DIRECTORES

Si el universo de interés es alguno de los siguientes:

- maestros urbanos de matemática
- maestros urbanos de lengua
- maestros rurales
- profesores de matemática
- profesores de lengua
- directores de primaria
- directores de secundaria

se debe tener presente que las respectivas muestras quedan determinadas por el procedimiento de selección de los alumnos. Así, maestros y profesores son seleccionados con probabilidades proporcionales a la cantidad de grupos del universo en los que dictan clase. Por su parte, los directores son seleccionados con probabilidad proporcional al número de secciones (divisiones) del establecimiento que dirigen.

A falta de información en relación con el número de grupos en los que dicta clase cada maestro o profesor, se supondrá que todos ellos, en el estrato al cual pertenecen, son seleccionados con probabilidades iguales. En consecuencia, el diseño muestral para estas unidades de observación es el muestreo aleatorio estratificado, en una sola etapa de selección, con probabilidades iguales. El estimador que se propone para la media de X es:

$$\tilde{\mu} = \sum_{h=1}^{K} \frac{M_h}{M} \tilde{\mu}_h, \text{ con } \tilde{\mu}_h = \frac{\sum_{j}^{m_h} X_{jh}}{m_h}$$

donde: X_{jh} es el valor de la variable X que toma la unidad observada (maestro o profesor) en el conglomerado j (sección o división) del estrato h.

La varianza de $\tilde{\mu}$ resulta igual a:

$$V(\tilde{\mu}) = \sum_{h=1}^{K} \left(\frac{M_h}{M}\right)^2 V(\tilde{\mu}_h) = \sum_{h=1}^{K} \left(\frac{M_h}{M}\right)^2 \frac{\sigma_h^2}{m_h} \frac{M_h - m_h}{M_h - 1}$$

donde σ_h^2 es la varianza de la variable X en el estrato h.

Como σ_h^2 es desconocida, se la estima a través de la propia muestra utilizando un estimador insesgado. De esta forma se obtiene un estimador para $V(\tilde{\mu})$:

$$\tilde{V}(\tilde{\mu}) = \sum_{h=1}^{K} \left(\frac{M_h}{M}\right)^2 \frac{1}{m_h} \frac{M_h - m_h}{M_h - 1} \frac{\sum_{j}^{m_h} (X_{jh} - \tilde{\mu}_h)^2}{m_h - 1}$$

En cuanto a los directores, tanto del nivel primario como del medio, el supuesto de que se han seleccionado con probabilidades iguales (dentro del estrato) es bastante fuerte, por cuanto el director de un colegio primario con 5 secciones tiene 5 veces más probabilidades de pertenecer a la muestra que el director de un establecimiento con una única sección en el 7º grado. En consecuencia, el diseño muestral para esta unidad de observación es el muestreo aleatorio estratificado, en una sola etapa de selección, con probabilidades desiguales.

El estimador de la media de X para la unidad de observación directores de un nivel sería:

$$\tilde{\mu} = \sum_{h=1}^{K} \frac{E_h}{E} \tilde{\mu}_h, \text{ con } \tilde{\mu}_h = \frac{1}{E_h} \sum_{j}^{m_h} \frac{X_{jh}}{\pi_{jh}}$$

donde: E_h = Nº de establecimientos del correspondiente nivel con al menos un conglomerado del universo de alumnos en el estrato h

π_{jh} = probabilidad de selección del establecimiento h en la muestra de m_h conglomerados del estrato h.

Como puede observarse, este estimador tropieza con el inconveniente que resulta del necesario conocimiento de π_{jh}, tarea difícil pero no imposible, y que consiste en acondicionar el marco muestral en términos de establecimientos escolares.

El cálculo de la varianza de $\tilde{\mu}$ plantea dificultades adicionales .**

Una alternativa para resolver el problema anterior consiste en la técnica de la estratificación a posteriori. Esta solución requiere, como en el caso anterior, acondicionar el marco muestral en téminos de establecimientos. En cada estrato (intersección de jurisdicción y ámbito o jurisdicción y modalidad) los establecimientos se clasificarán en función del número de conglomerados (secciones o divisiones) de alumnos. Sea $E_{(t)h}$ = número de establecimientos con t conglomerados en el universo en el estrato h (t = 1,2,3,.....T).

En la muestra, en el estrato h, los directores se clasifican en clases, en función del número de conglomerados de su establecimiento.

Sean:

$m_{(t)h}$ = número de conglomerados en la muestra del estrato h pertenecientes a establecimientos con t conglomerados.

** Ver por ejemplo Sanchez-Crespo "Curso intensivo de muestreo en poblaciones finitas" Cap. X.

$X_{jh}^{(t)}$ = valor de la variable X en el conglomerado j de la muestra en el estrato h, donde el conglomerado j pertenece a un establecimiento con t conglomerados.

$$E_{(t)} = \sum_{h=1}^{K} E_{(t)h}$$

$$E = \sum_{t=1}^{T} E_{(t)}$$

Entonces, el estimador de la media de X para los directores de un nivel, sería:

$$\tilde{\mu} = \sum_{t=1}^{T} \frac{E_{(t)}}{E} \tilde{\mu}_{(t)}, \quad \text{donde } \tilde{\mu}_{(t)} = \frac{\sum_{h=1}^{K} \sum_{j=1}^{m_{(t)h}} X_{jh}^{(t)}}{\sum_{h=1}^{K} m_{(t)h}}$$

$\tilde{\mu}_{(t)}$ es un estimador de razón combinado a través de los anteriores estratos (h). $\tilde{\mu}_{(t)}$ podría también obtenerse como promedio ponderado de estimadores de razón separados, uno por cada estrato a posteriori, pero ello no es recomendable en este caso en virtud del reducido tamaño de la muestra en cada estrato, con lo cual en muchos casos podría conducir a denominadores nulos ($m_{(t)h} = 0$ para algunos valores de t).

Sea $m_{(t)} = \sum_{h=1}^{K} m_{(t)h}$. Entonces un estimador apropiado para $V(\mu)$ es:

$$V(\mu) = \sum_{t=1}^{T} \left(\frac{E_{(t)}}{E} \right)^2 \frac{s_{(t)}^2}{m_{(t)}} \frac{E_{(t)} - m_{(t)}}{E_{(t)} - 1}$$

donde $$s^2_{(t)} = \frac{\sum_{h,j} (X^{(t)}_{jh} - \mu_{(t)})^2}{m_{(t)} - 1}$$

9. CONSIDERACIONES FINALES

9.1 IDENTIFICACION DE LOS FORMULARIOS

Se deduce del plan de estimadores que es esencial para los cálculos la correcta asignación de cada formulario de alumnos al correspondiente conglomerado, ámbito o modalidad, y jurisdicción. En consecuencia, estos datos de identificación deberán ingresarse al computador como un atributo principal del alumno.

A los efectos del análisis de correlación entre variables, es necesario verificar en el ingreso de datos que las identificaciones de los diferentes formularios de un mismo alumno sean coincidentes (prueba de matemática, prueba de lengua, cuestionario del alumno, cuestionario de la familia en el nivel primario).

Finalmente, para correlacionar variables del alumno con los del maestro o profesor, es necesario que la identificación de estos coincida con la de la sección o división de sus alumnos.

Un cuidado especial debe observarse con la identificación de los cuestionarios en ocasión de la correcciones que será necesario realizar como consecuencia de la crítica y detección de inconsistencias.

9.2 ADVERTENCIAS

La muestra no es autoponderada; esto es, los resultados de la muestra no pueden agregarse simplemente (por ejemplo para estimar promedios o proporciones), sino que previamente deben clasificarse por conglomerado, ámbito o modalidad, y jurisdicción, y luego ponderarse adecuadamente para obtener los estimadores tal como se comentó en la sección 8.

En la determinación del tamaño de la muestra se ha trabajado con el mínimo necesario para obtener cierta precisión y seguridad. Si no se pone especial cuidado en el sistema de archivo de los cuestionarios (que podrán alcanzar un orden de magnitud de 100.000), aún cuando la muestra se hubiera realizado en el 100% en el campo, una

parte de ella podrá "perderse" en las siguientes etapas (archivo - codificación - procesamiento electrónico - crítica - levantamiento de inconsistencias - concentración) y aunque la parte perdida sea pequeña, no hay que esperar que la pérdida se distribuya proporcionalmente a lo largo de la muestra. Antes bien, es probable que la información perdida se concentre en alguna de las jurisdicciones y que ello impida obtener inferencia válida a ese nivel. Entonces, el archivo debe ser dotado de un sistema de seguridad, un jefe de alto nivel y un equipo de computación que a través de un programa adecuado permita conocer, a cada instante, en qué etapa del procesamiento se encuentra cada lote de formularios.

Una disminución en el tamaño de la muestra en un estrato puede ocasionar una pérdida importante de precisión. Téngase presente que, para todos los estratos, es imprescindible que $m_h \geq 2$, pues de lo contrario no será posible estimar varianzas.

9.3 CALCULO DE LA PRECISION A POSTERIORI

A los efectos de conocer la bondad de las estimaciones puntuales, se sugiere realizar el cálculo de la precisión (absoluta), y verificar, para el caso particular de las pruebas de lengua y matemática, si los tamaños de muestra resultaron suficientes para lograr la precisión deseada con un 95% de seguridad. Cualquiera sea la variable X considerada, si $\mu = E(X)$ es el parámetro que se quiere estimar, para una seguridad del 95%, la precisión es ϵ tal que:

$$P(| \tilde{\mu} - \mu | < \epsilon) = 0.95$$

Si el tamaño de la muestra es "grande", aplicando los teoremas de convergencia se tiene que:

$$\tilde{\epsilon} = 1.96 \sqrt{\tilde{V}(\tilde{\mu})}$$

donde $\tilde{V}(\tilde{\mu})$ se calcula, en cada caso, de acuerdo con las fórmulas de la sección 8.

Obsérvese que, adicionalmente, 2ϵ es la amplitud de un intervalo de confianza para μ, con lo que al calcular $\tilde{\epsilon}$ se podrá construir un intervalo para μ.

9.4 ESTIMACION DE LA CORRELACION INTRA-CONGLOMERADO

Para la determinación del tamaño de la muestra de alumnos, en cada estrato, se supuso que la correlación intra-conglomerado (ρ_h) era positiva y mayor cuanto menor el tamaño del conglomerado (0.10 para $T_h = 25$ y 0.5 para $T_h = 6$). Este supuesto tuvo como consecuencia un incremento de la varianza de 3,5 veces en el muestreo por conglomerados respecto del muestreo aleatorio simple. Probablemente este supuesto exagera el efecto "contagio". Si así fuera, los tamaños de muestra calculados para 1993 habrían sido sobreestimados. Para verificar este extremo, y poder así mejorar el cálculo del tamaño de la muestra para 1994, se propone estimar el coeficiente de correlación intraconglomerado, a través de un procedimiento simplificado.

Utilizando la misma notación de las secciones anteriores se tiene que

$\tilde{V}(\tilde{\mu}_h)$ es la varianza del estimador de μ_h en el muestreo por conglomerados. Considérese ahora la muestra de alumnos del estrato h. Su tamaño es:

$$n_h = \sum_{j=1}^{m_h} T_{jh}$$

Supóngase que estos n_h alumnos provienen de una muestra aleatoria simple sin reposición. Entonces, el estimador de $V_{MAS}(\tilde{\mu}_h)$ es

$$\tilde{V}_{MAS}(\tilde{\mu}_h) = \left(1 - \frac{n_h}{N_h}\right)\frac{S_h^2}{n_h}$$

donde $S_h^2 = \dfrac{\sum_{i=1}^{n_h} (x_{ih} - \bar{x}_h)^2}{n_h - 1}$ y $\bar{x}_h = \dfrac{\sum_{i=1}^{n_h} x_{ih}}{n_h}$

El cociente de ambas varianzas, $\tilde{V}(\tilde{\mu}_h)$ sobre $V_{MAS}(\tilde{\mu}_h)$, es el "efecto de diseño" de KISH: $[1 + (T_h - 1)\rho_h]$. Entonces, utilizando las respectivas estimaciones, es posible despejar $\tilde{\rho}_h$ a partir de la ecuación siguiente.

$$\frac{\tilde{V}(\tilde{\mu}_h)}{\tilde{V}_{MAS}(\tilde{\mu}_h)} = 1 + (\tilde{T}_h - 1)\,\tilde{\rho}_h$$

En el caso particular en que X es el resultado de las pruebas, es conveniente evaluar en qué medida el número 3,5 (sección 7, página 12) resultó una buena aproximación del efecto de diseño.

9.5 ESTIMACION DE VARIANZAS PARA OPERATIVOS POSTERIORES

El procedimiento recomendado en 9.4 proporciona, lateralmente, un procedimiento para estimar las varianzas del rendimiento escolar en las pruebas de cada nivel, en cada estrato. Las σ_h^2 pueden aproximarse razonablemente por las S_h^2 según se definen en 9.4. De esta manera se contará con un banco de varianzas por estrato, cuyos valores determinarán para 1994 en adelante, la conveniencia (o no) de optar por un diseño estratificado óptimo, atendiendo a la variabilidad de la variable X y a los costos del operativo en cada estrato.

****

ANEXO 1: ESTIMADORES ALTERNATIVOS PARA μ y $V(\tilde{\mu})$

Si se desea confrontar las estimaciones de razón (sección 8) con otra alternativa, se presentan aquí los estimadores de "expansión simple" o clásicos para μ, donde solo es aleatorio el numerador. Utilizando la misma notación que en la sección 8, se tiene:

$$\tilde{\mu} = \sum_{h=1}^{K} \frac{N_h}{N} \hat{\mu}_h, \text{ con } \hat{\mu}_h = \frac{\sum_{j}^{m_h} \sum_{i}^{T_{jh}} X_{ijh}}{m_h T_h}$$

Entonces: $V(\tilde{\mu}) = \sum_{h=1}^{K} \left(\frac{N_h}{N}\right)^2 V(\hat{\mu}_h)$, donde el problema consiste en encontrar $V(\hat{\mu}_h)$. Una aproximación del problema se obtiene a partir de (W. Cochran Op. cit.):

$$V(\hat{\mu}_h) \cong \left(1-\frac{m_h}{M_h}\right) \frac{1}{m_h T_h^2} \frac{\sum_{j}^{M_h} T_{jh}^2 \left(\bar{X}_{jh} - \frac{1}{M_h} \sum_{j}^{M_h} \bar{X}_{jh}\right)^2}{M_h - 1}$$

Una estimación de $V(\hat{\mu}_h)$ puede obtenerse utilizando:

$$\tilde{V}(\hat{\mu}_h) = \left(1-\frac{m_h}{M_h}\right) \frac{1}{m_h \tilde{T}_h^2} \frac{\sum_{j}^{m_h} T_{jh}^2 \left(\bar{x}_{jh} - \frac{1}{m_h} \sum_{j}^{m_h} \bar{x}_{jh}\right)^2}{m_h - 1}$$

A N E X O 2

Consideraciones sobre los códigos de identificación de los cuestionarios.
Notas para el procesamiento y el plan de tabulados.

CUESTION Indica el código que corresponde a cada uno de los 17 formularios a utilizarse en el relevamiento.
- Cuando CUESTION = 1 ó 2, es necesario especificar AMBITO
- Cuando CUESTION = 12,13,14,15,16,17, es necesario especificar MODAL
- En los restantes casos el código CUESTION permite generar automáticamente el AMBITO:

* Si CUESTION = 3,5,7,8 ó 10 ⇒ AMBITO = 1 (urbano)
* Si CUESTION = 4,6,9 u 11 ⇒ AMBITO = 2 (rural)

Máscaras

	P					
	C	J	A	S	Ó	A
	n	C	o			
P	0					
P	0					
C	0		▒			
C	0		▒			
C	0		▒			
C	0		▒			
C	0		▒			
C	0		▒			
C	0		▒			
C	1		▒			
C	1		▒			

	M					
	C	J	M	S	Ó	A
	.	.	l			
P	1					
P	1					
C	1					
C	1					
C	1					
C	1					