

1. Introducción.

En este informe se discuten las afirmaciones hechas en la denuncia presentada por el Ing. Elices contra las empresas CIFRA, Equipos Mori y Factum, lo expresado en las defensas de las empresas y en la contestación de Elices a las defensas. A los efectos de recorrer los distintos puntos se ha dispuesto el informe en varias secciones. La Sección 2 revisa los cálculos hechos por Elices y por un investigador de la Universidad de Michigan contactado por Elices, el Dr. Michael Elliot. En la Sección 3 se analizan las respuestas dadas por las empresas a la denuncia. En la Sección 4 se estudia la respuesta a las empresas del Ing. Elices y en la Sección 5 se tratan de responder las preguntas dirigidas al “experto actuante” en este expediente.

La discusión de los distintos aspectos no me permite llevar un “hilo conductor”, por lo cual en la Sección 6 se resumen las principales conclusiones respecto a la denuncia.

2. Cálculos iniciales, según escritos presentados por Ing. Elices.

En sus escritos de fecha 26 de noviembre de 2009 (fojas 1-9 del expediente) y 14 de diciembre de 2009 (fojas 13 y sgtes.) el Ing. Francisco Elices presenta cálculos de probabilidades realizados por él mismo y por el Dr. Michael Elliot del Departamento de Bioestadística de la Universidad de Michigan. Estos cálculos conducen a probabilidades muy pequeñas, que hacen pensar en la no independencia entre las estimaciones de las empresas denunciadas.

Hay dos puntos a remarcar respecto a estos cálculos:

- Efecto de redondeo. Las empresas publican sus resultados con dos cifras decimales, o dicho de otra manera, en números enteros al referirse a porcentajes (en uno de los meses, una de las empresas publica su resultado con tres cifras significativas)
- Hay distintas formas de medir similitudes o diferencias en una situación dada. El Ing. Elices elige medir la máxima diferencia en cada uno de los meses por separado mientras que el Dr. Elliot calcula la variabilidad total de los nueve valores disponibles: 3 meses x 3 empresas por mes¹. Estas distintas formas de medir diferencias pueden conducir, y conducen, a resultados diferentes.

Efecto de redondeo: resultados discretos y no continuos.

Las tres empresas denunciadas usan tamaños muestrales cercanos a 1.000, por lo cual la distribución del estimador del porcentaje de votantes a un partido puede aproximarse muy bien con una distribución normal con media p y la varianza que dependa del diseño muestral, siendo p el verdadero porcentaje

¹ Al momento de consultarlo al Dr. Elliot se le enviaron resultados de los meses julio, agosto y setiembre

poblacional de ese partido². Tanto Elices como Elliot toman la varianza como si fuese la del muestreo aleatorio simple: $p(1 - p)/1000$.

La distribución normal corresponde a una variable aleatoria continua, mientras que los resultados publicados son discretos, con dos cifras significativas en general. Lo que hacen tanto el Ing. Elices como el Dr. Elliot, y lo que se hace en los cálculos que se presentan más adelante, es considerar variables aleatorias normales con media p y varianza $p(1 - p)/1000$ y redondear el resultado a dos cifras significativas. Las variables resultantes del redondeo son entonces discretas, sólo pueden tomar un número finito de valores, y lo mismo ocurre con las funciones u operaciones que las involucren.

Una estimación de 46% se debe a un resultado real entre 45.50% y 46.49% o, para ser más precisos, a un resultado en el intervalo semi-cerrado por la izquierda [45.5%, 46.5%). El efecto de “discretización” es aún más destacado en las diferencias entre estimaciones: 45.50% y 46.49% están a una distancia de 0.00% (son iguales a los efectos observables) mientras que 45.49% y 45.50% están a una distancia de 1.00%.

Este efecto de “discretización” es comentado en el correo electrónico de Elliot (foja 9 del expediente) cuando prefiere usar una solución alternativa a la inicialmente sugerida por el Dr. Rod Little, su colega de la Universidad de Michigan.

Medidas de similitud, distintas opciones.

En general todos podemos consensuar en lo que significa que dos o más cosas sean “iguales”, pero no necesariamente sea el caso cuando decimos que dos o más cosas son “similares” o “parecidas”. La base de la denuncia del Ing. Elices es que los resultados observados de las tres empresas encuestadoras “son tan similares entre sí” que no pueden deberse a observaciones independientes. Para medir la similitud, o mejor dicho la disimilitud, entre los resultados el Ing. Elices selecciona medir la distancia entre la menor y la mayor de las estimaciones en *cada* uno de los cuatro meses julio-octubre, mientras que el

² Para los tamaños muestrales usados, la aproximación normal es buena si el parámetro p no está cerca de los extremos 0 o 1, tal como es el presente caso.

Dr. Elliot calcula la varianza muestral entre las nueve observaciones correspondientes a los resultados publicados por las tres empresas en los tres meses de julio a setiembre³. Esta dualidad al elegir medidas de disimilitud lleva a resultados que no coinciden, lo cual no es de sorprender si consideramos que usan medidas y metodologías diferentes, uno mide diferencias en cada mes, mientras que el otro considera los tres meses conjuntamente.

Cálculos del Dr. Michael Elliot.

El Dr. Elliot usa un modelo que considera las observaciones de los tres meses y las tres empresas como siendo generadas independientemente por una misma distribución poblacional. Esto implica que las tres empresas toman muestras de una misma población, que esa población puede suponerse incambiada en el tiempo, el porcentaje p permanece constante en los tres meses y las empresas operan independientemente entre ellas y consigo mismas a través del tiempo. A partir de estos supuestos, se aproxima la distribución del estadístico elegido para medir disimilitudes mediante simulación Monte Carlo.

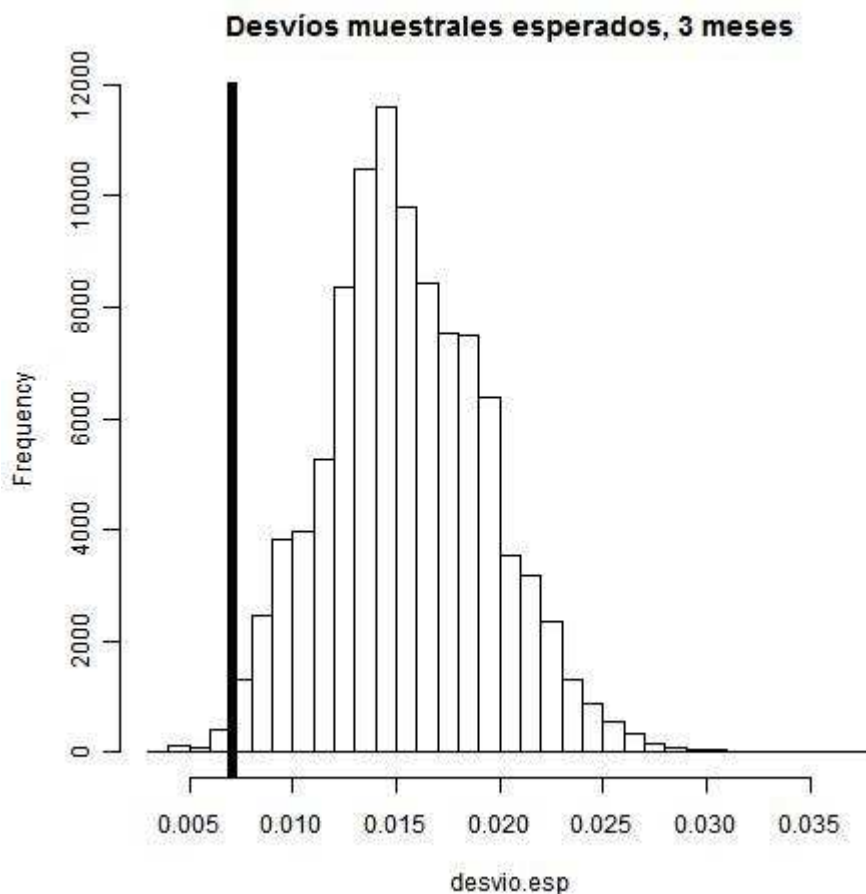
Elliot mide la disimilaridad mediante la varianza muestral de los nueve resultados que dispone. La varianza muestral observada es de 0.00005 y Elliot encuentra que la probabilidad de observar varianzas menores a ese valor es del orden de 0.006 (ver correo electrónico en foja 9).

Se trató de replicar el cálculo de Elliot usando simulación Monte Carlo seleccionando un millón de muestras independientes de tamaño 9 de una normal con media $\hat{p} = 0.4466667$, el promedio de los nueve resultados, y varianza $\hat{p}(1 - \hat{p})/1000$. La probabilidad de encontrar desvíos *menores* que el observado, o lo que es lo mismo, varianzas menores que la observada es del orden de 0.006⁴. Este resultado coincide con el presentado por Elliot (seis en 1000). Bajo los supuestos del modelo, resultados como los observados sólo se observarían menos del 1% de las veces. Si bien la presente podría ser

³ El correo electrónico del Dr. Elliot está fechado el 12 de octubre, cuando aún no estaba disponible la estimación correspondiente a ese mes.

⁴ En lo que refiere a probabilidades estimadas o aproximadas, se presentan los resultados con el número de dígitos significativos al 99.99% de confianza, ver Apéndice 3.

una de esas seis de cada mil veces que esto puede ocurrir por casualidad, la probabilidad calculada es lo suficientemente pequeña para despertar sospechas sobre la independencia entre las empresas, la independencia entre los distintos meses, o ambas cosas a la vez. Se repitió la misma metodología usando los doce resultados (que Elliot no llegó a disponer) y la probabilidad es de 0.008⁵. Cuando se usan los doce meses la estimación del porcentaje de votos al Frente Amplio es de $\hat{p} = 0.4495833$. El programa en R usado para replicar el cálculo de Elliot para 3 meses, y el equivalente para los 4 meses se adjunta en el Apéndice 1.



En el gráfico anterior se presenta una aproximación a la distribución en el muestreo del desvío estándar muestral. La línea recta vertical en 0.007 (el desvío correspondiente a una varianza de 0.00005) muestra que sólo una proporción muy pequeña de desvíos son menores o iguales al observado.

⁵ Si se consideran desvíos *menores o iguales* al observado se obtienen probabilidades similares a las calculadas: 0.011 y 0.008 para 9 y 12 pronósticos respectivamente.

Cálculos del Ing. Francisco Elices.

El Ing. Elices usa un modelo similar al de Elliot para generar las observaciones de los cuatro meses en forma independiente y provenientes de una misma distribución normal con media “aproximadamente el 45%” (foja 14). Se tomó el promedio de las tres empresas para los cuatro meses $\hat{p} = 0.4495833$ y varianza igual a $\hat{p}(1 - \hat{p})/1000$. La medida de disimilaridad usada por Elices es la distancia máxima entre pronósticos para *cada* mes. En los datos observados esa distancia es siempre menor o igual a 1%.

La medida de disimilitud usada por Elices, máxima distancia entre las proyecciones es muy sensible a la discretización de los resultados. La distancia máxima entre resultados para un mismo mes sólo puede tomar los valores 0%, 1%, 2%, 3%, etc., por lo que resulta *muy* importante distinguir entre diferencias *menores* o diferencias *menores o iguales* a las observadas. Los cálculos del Ing. Elices están hecho considerando diferencias *menores o iguales* a 1%.

El programa en R que aparece en el Apéndice 2 calcula la probabilidad aproximada de que en un mes la diferencia máxima entre pronósticos sea menor o igual a 1%. Se consideraron los datos de los cuatro meses, y se generaron doce millones de observaciones normales con media $\hat{p} = 0.4495833$ y varianza $\hat{p}(1 - \hat{p})/1000$, esto es, el equivalente a un millón de muestras de tamaño 12. La probabilidad de que para un mes cualquiera la diferencia máxima entre pronósticos sea menor o igual a 1% es 0.225. Si consideramos que esta diferencia nunca sobrepasó 1% durante cuatro meses, el Ing. Elices propone que dicha probabilidad es de $0.225^4=0.0025$, aproximadamente un 2.5 por mil. Nótese que para llegar a esa probabilidad estamos aceptando el supuesto de independencia entre meses para cada una de las empresas.

Si consideramos, al igual que hizo Elliot, que las diferencias máximas sean estrictamente *menores* a 1%, entonces la probabilidad de que en un mes se encuentren resultados al azar donde la diferencia sea 0% baja de 0.225 a 0.036, y a su vez, para que la diferencia sea menor a 1% durante los cuatro meses considerados, la probabilidad baja de 2.5 en mil a aproximadamente 1 en 600.000 (0.00000167).

3. Respuestas de las empresas demandadas.

En el mes de enero de 2010 las empresas demandadas presentan escritos donde contestan la demanda del Ing. Elices.

Equipos MORI.

La respuesta de Equipos MORI a la demanda figura a fojas 75 y sgtes. del expediente firmada por César A. Aguiar.

Algunos comentarios al respecto son:

- Aguiar observa (puntos 2 y 3) que Elices elige tres de las empresas, pero que en la plaza hay más empresas que se dedican a realizar encuestas de intención de voto. Asimismo, si bien hubo muchas mediciones entre febrero y noviembre, Aguiar apunta que Elices “selecciona ... aquellas que pueden fundar sus argumentos”. Es cierto que elegir “arbitrariamente” empresas y/o instancias de comparación debería ser considerado al calcular las probabilidades de tan poca disimilaridad. Si uno elige 3 de 5 empresas, está seleccionando una de las diez posibles combinaciones de tres empresas seleccionadas de un grupo de cinco, por lo que las probabilidades tan pequeñas calculadas en la sección anterior deberían ser aumentadas en un orden de magnitud ($C_3^5 = 10$).
- Aguiar plantea que los márgenes de error pueden ser menores a los publicados por varias razones (puntos 6 y 7):
 - las empresas disminuyen la “varianza de resultados” luego de las elecciones internas “por la mejora en los procedimientos de ponderación” (punto 6), es decir, además de los datos de la encuesta de ese mes, tienen los resultados de las elecciones internas y los resultados de las encuestas de meses anteriores
 - las empresas disponen de la información de las otras empresas, la “coordinación implícita, ex – post, no necesariamente

conciente” relacionada con la información publicada por las empresas de la competencia (punto 7).

- De acuerdo a Aguiar, con el cúmulo de información propia y publicada por otros, las empresas “proceden a ponderar sus bases empíricas antes de difundir sus nuevos resultados finales”, esto es, lo *publicado no necesariamente coincide* con el resultado empírico de la encuesta. Esta separación entre “encuestas de intención de voto” y un preceso posterior de “estimación de voto” se repite, con diferentes denominaciones, en las defensas de las tres empresas y en la respuesta a las defensas de Elices.

CIFRA.

La respuesta de CIFRA a la demanda figura a fojas 87 y sgtes. del expediente firmada por Luis E. González.

Al respecto podemos comentar:

- González no comparte que en la demanda se elija solamente el grado de acierto al Frente Amplio, cuando “... la calidad de las estimaciones preelectorales se mide teniendo en cuenta, para *todos* los partidos que participan en la elección...” (punto 6).
- Los resultados de las encuestas se someten a procesamientos posteriores, pero “las consideraciones del Ing. Elices ... se aplican luego a resultados que dependen de la aplicación de otros procedimientos no probabilísticos” (punto 9). González dice que “las estimaciones de las encuestadoras también tienen en cuenta otros datos a través de métodos de ponderación, ajuste, post estratificación o similares que ese apoyan en unos pocos modelos teóricos, que analizan por qué razón los resultados directos de las encuestas deben ser revisados” (punto 10), por lo cual los cálculos probabilísticos y las técnicas del Ing. Elices “sólo podrían aplicarse (eventualmente) a los resultados directos de las encuestas *antes* de su ponderación, ajuste o post estratificación” (punto 11).

- González apunta que en algunos artículos publicados “se analiza por qué tienden a converger los resultados de las encuestas⁶ a medida que se aproxima la fecha de la elección” (punto 12). En *Survey Practice* aparece un artículo (<http://surveypractice.org/2008/12/19/the-experts-reply-convergence-mystery/>) donde varios expertos dan razones similares a las de González: procesamientos *posteriores* a las encuestas son los que explican la disminución en la variabilidad de los resultados de distintas empresas a medida que se acerca la fecha de una elección.

Factum.

La respuesta de Factum a la demanda figura a fojas 91 y sgtes. del expediente. Está firmada por Oscar Alberto Bottinelli.

Entre los puntos que contesta Bottinelli figuran:

- Bottinelli afirma que las encuestas de opinión pública son “investigaciones científicas en ciencias sociales de tipo cuantitativo” y como tal deben ser analizadas o discutidas (punto 1.1.3).
- Bottinelli dice que “solamente una etapa de la muestra es de carácter aleatorio”, pero que las encuestas no son juegos de azar. (punto 1.1.5)
- En el punto 1.1.6 Bottinelli sostiene que porque algo tenga probabilidad baja no es suficiente para despertar sospechas. En particular se plantea un ejemplo, “si en un juego de ruleta que el número “18” salga diez veces consecutivas puede resultar llamativo” eso no debería dar lugar a “ningún procesamiento, indagación policial ni sumario administrativo”. No estoy de acuerdo con el planteo, a pesar de mi poca experiencia en casinos, la probabilidad de que el 18 salga diez veces seguidas es *tan* pequeña que indudablemente va a dar lugar a que se investigue, lo que en definitiva es la base de la acusación del Ing. Elices. Si en una ruleta con 37 números se juega una vez por minuto día y noche, todos los días del año, que el 18 salga diez veces consecutivas ocurriría en promedio una vez cada *ciento cincuenta millones de años*. Supongo que cualquier

⁶ En la misma lógica de González, supongo que sería más apropiado decir que “tienden a converger las” *predicciones electorales* que es lo que se divulga, y no las encuestas en sí.

encargado de una sala ciertamente va a desconfiar (y a investigar) si algo así ocurre.

- En el punto 1.2.10 se plantea que se deben “analizar e interpretar correctamente los datos, y difundirlos a los clientes o a la opinión pública”. Aún cuando no está expresado tan claramente como en las respuestas de Equipos MORI y de CIFRA, entiendo que las cifras que se difunden no son el resultado de las encuestas, sino la interpretación (o manipulación, en términos de Elices) de los datos concretos que hacen “los responsables de la institución”.

4. Contestación del Ing. Elices a las empresas.

En un correo electrónico de fecha 25 de enero de 2010 el Ing. Elices agrega información adicional sobre la “baja dispersión de los resultados” que ocurrió con motivo de la publicación de pronósticos para el balotaje durante tres semanas de noviembre. Se presenta en dicho correo información para la semana del 2 al 8 de noviembre (tres empresas), para la semana del 11 al 15 de noviembre (dos empresas) y para la semana del 19 al 22 de noviembre (tres empresas). Si actualizamos el cálculo del Ing. Elices, ahora considerando que la distancia entre las empresas permanece siendo menor o igual a 1% y consideramos que esto ocurrió en seis instancias (considero los cuatro meses julio-octubre de la denuncia original, más dos de las tres semanas de noviembre), la probabilidad de ocurrencia de una diferencia tan pequeña sólo debido al azar sería del orden de $0.225^6=0.00013$.

Respuesta de Elices.

Con fecha 8 de marzo de 2010 se presenta la respuesta del Ing. Elices a las defensas de las empresas (fojas 123 y sgtes del expediente). La respuesta consta de una nota (págs. 1 y 2), una ampliación sobre los puntos por él planteados y respuestas a los argumentos de las empresas (págs. 3 a 21) y diversos documentos adjuntos (págs. 22 y siguientes del escrito de Elices).

Algunos comentarios al respecto:

- El Ing. Elices comparte que la realización de encuestas requiere “del concurso de otras disciplinas como la sociología y la ciencia política”, pero que el punto clave de estos procedimientos es determinar si hubo o no manipulación de datos y, en caso de haberla, si eso afecta o no la libre competencia (pág. 1 de respuesta de Elice).
- Debemos determinar si la “extraordinaria concordancia es resultado de una manipulación, ajuste o retoque de los resultados de las encuestas, o

por el contrario, es producto del azar” (pág 2). Estoy de acuerdo con Elices, este es el punto clave, y más adelante me extenderé al respecto.

- Elices considera una especie de confesión de parte la “explicación alternativa” de Aguiar respecto de que hay una *coordinación implícita* entre las empresas que lleva a los encuestadores a *atender* los resultados ya publicados por sus colegas cuando ponderan sus bases empíricas antes de difundir los resultados finales (punto 7 de la defensa de Equipos) (pág. 2).
- Elices remarca que “... la ponderación de bases de encuestas independientes no se puede hacer, por definición, en función de los resultados de otras encuestas”. En caso de hacerlo, la encuesta así ponderada deja de ser independiente (pág 2). Es cierto, para que las encuestas sean independientes deben ser realizadas y procesadas independientemente.
- **“Resultados de la encuesta de intención de voto” vs. “estimación del voto”**. En la pág. 6 del escrito de Elices, en la sección “3) Aclaración de la naturaleza de la denuncia”, el Ing. Elices plantea una distinción entre dos productos de las empresas encuestadoras⁷,
 1. por un lado el **resultado de la encuesta de intención de voto**, los tabulados correspondientes al procesamiento computacional de la encuesta, tabulados realizados con una “metodología precisa”. Básicamente, estamos hablando de lo que resulta del procesamiento mediante un paquete estadístico de los microdatos muestrales.
 2. Por otro, la **estimación del voto**, o estimación del voto real, proceso *posterior* al descrito en el punto anterior que implica “como mínimo la asignación de los ‘indecisos’ a alguno de los partidos o al ‘voto en blanco. También puede realizar otras consideraciones y modificaciones [... el encuestador ...] de

⁷ Como ya se dijo, estos dos productos aparecen referenciados bajo distintos nombres en las defensas de las tres empresas.

acuerdo a su mejor criterio y experiencia”. En este segundo *procesamiento* de la información proveniente de los “resultados de la encuesta” se utilizan metodologías heurísticas y este “proceso sí requiere ‘la intervención de la mano humana’ y en este caso no es una manipulación de datos, sino un proceso discrecional válido y donde los aportes de la Sociología y la Ciencia Política pueden resultar valiosos”.

3. Elices plantea que el error muestral, entendido como medida de variabilidad en los resultados, debe mantenerse en la segunda etapa (estimación del voto) respecto de la primera (encuesta), y que este error no se puede reducir. En mi opinión, la segunda etapa no se rige por las reglas del muestreo probabilístico, y por lo tanto no se pueden aplicar a las estimaciones de voto conceptos como varianza, insesgamiento, etc., puesto que hay otras disciplinas que toman los resultados de la encuesta como *input* y producen las estimaciones de voto como *output*. El proceso que la Sociología o la Ciencia Política realizan tomando como uno de sus insumos los resultados de la encuesta de intención de voto *no* se rige por conceptos estadísticos o probabilísticos y por lo tanto no pueden aplicarse cálculos sobre el error muestral.
4. Al final de la página 6 está resaltada la siguiente frase: “La denuncia por mí presentada, se refiere únicamente a las encuestas de intención de voto, y no hace ninguna referencia a los trabajos de estimación de voto”. El problema es que lo divulgado por las empresas mes a mes, lo que aparece publicado, y en definitiva, los datos en los que se basa la denuncia, son las estimaciones de voto y *no* los resultados de las encuestas. Las tres empresas describen que hay una etapa posterior a los tabulados de las encuestas, y el mismo Ing. Elices lo acepta en la pág. 6 de su respuesta a las empresas, citada en el punto 2 previo:

- i. Aguiar lo dice en su punto 7 (foja 79) que las empresas “proceden a ponderar sus bases empíricas antes de difundir sus nuevos resultados finales”.
 - ii. El proceso de “anclaje” mencionado por González en los puntos 9 y 10 (foja 89).
 - iii. La etapa de “analizar e interpretar correctamente los datos, y difundirlos a los clientes o a la opinión pública” que Bottinelli plantea en su punto 1.2.10 (foja 97).
- Elices en la pág. 9 de su escrito plantea el tema de la “razonabilidad” de la denuncia en respuesta al planteo de Aguiar de que no hay otras pruebas más allá de las estadísticas. Comparto en ese sentido las razones que llevaron a Elices a presentar la denuncia puesto que los resultados, desde un punto de vista probabilístico, aparecen como demasiado similares. Elices cita a Marabotto en que “la prueba debe ser valorada racionalmente de acuerdo con las reglas de la sana crítica” (pág 9).
 - En la pág. 13 y sgtes. se discute el correcto cálculo del “error estadístico” publicado por las empresas. Los diseños utilizados en las etapas que llevan al cálculo de los “resultados de la encuesta” son en general lo bastante complejos como para que no pueda hacerse el cálculo exacto del error muestral. Además de usar estratificación, conglomerados, múltiples etapas, etc. en estos diseños, hay aspectos de imputación de la no respuesta que hacen difícil el cálculo *correcto* de la varianza muestral y por lo tanto del error muestral o “error estadístico”. Ante la presencia de no-respuesta debemos usar modelos apropiados de cálculo de varianza⁸. Las tres empresas denunciadas publican como aproximación al error estadístico el que surge de una muestra aleatoria simple con una proporción poblacional igual (o cercana) a 0.50 para muestras de tamaño muestral cercano a 1000. Es ese error muestral

⁸ Ver por ejemplo “Model Assisted Survey Sampling”, Carl-Erik Särndal, Bengt Swensson y Jan Wretman (2003).

publicado el que Elices considera para sus cálculos. El verdadero error en los resultados de las encuestas de intención de voto puede ser *mayor* o *menor* al publicado por las empresas y al usado por Elices en sus cálculos, pero probablemente sea del orden del 3% tal como se publica en algunas de las fichas técnicas de las encuestas. El efecto diseño es muy fuerte en algunos casos, pero no parece serlo en las preguntas sobre intención de voto y en los diseños utilizados por las empresas. Si el diseño incluyese consultar a todos los residentes en condiciones de votar de un hogar, claramente sus respuestas no pueden considerarse como variables aleatorias independientes y el efecto conglomerado es muy importante en cuanto a incrementar la varianza, pero este caso no se da en el escenario que describimos pues las fichas técnicas de las empresas declaran que se selecciona un solo individuo por hogar. Los diseños sí consideran agrupaciones geográficas, y a ese respecto, si preguntamos a varios residentes del barrio La Teja en Montevideo obtenemos respuestas similares, pero en materia de preferencias políticas la edad de los entrevistados juega un papel muy importante, y los conglomerados geográficos que en general tienen los diseños tienden a cubrir individuos de edades variadas⁹.

- En las págs. 15 y sgtes. de su escrito (punto 4.5) el Ing. Elices reitera que la explicación de Aguiar que figura en el punto 7 de su defensa (fojas 78 y 79 del expediente) constituye “prácticamente una confesión de las prácticas denunciadas” y con ello cierra su línea argumental.

⁹ Un ejemplo del impacto de los conglomerados fue presentado en el VII Congreso Latinoamericano de Sociedades de Estadística en Rosario, Argentina (2006): “¿Cuántas personas vemos cuando vemos 409 personas?” Goyeneche, J.J., Urrestarazú, I. y Zoppolo, G.

5. Contestación a las preguntas del Ing. Elices.

Con fecha 5 de mayo de 2010 el Ing. Elices presenta un escrito que incluye varias consultas dirigidas al “experto actuante” (fojas 143 y sgtes. del expediente). Las respuestas u observaciones a las mismas se presentan a continuación.

1. Experimento mental del agua con hielo. Tal como dice el Ing. Elices, hay dos respuestas válidas, pero a mi entender las dos respuestas corresponden a campos diferentes:
 - a. Punto de vista teórico. Corresponde a considerar “un tiempo infinito, o universos infinitos” como plantea Elices en la pág. 3 de su escrito. Hay un ejemplo clásico en Teoría de la Probabilidad para tratar el concepto de infinito: se pone un mono frente a una máquina de escribir, durante un tiempo infinito el mono será capaz de escribir las obras completas de Shakespeare sin errores y sin omisiones, ... *infinitas* veces¹⁰. Desde este punto de vista no es imposible que los cubos de hielo hayan terminado en las copas sin la ayuda de las personas de la mesa ya sea porque aprendieron a volar por sí mismos, porque un meteorito los evaporó haciéndolos desaparecer mientras que otro meteorito de hielo caía en las copas o el escenario más inverosímil (e increíble) que se nos pueda ocurrir. Si asumo como posible, aunque improbable, que “o macaco” escriba las obras de Shakespeare, no puedo descartar la hipótesis de los hielos voladores o de que el 18 salga diez veces seguidas en la ruleta.
 - b. Punto de vista práctico o, yo agregaría, más razonable. A todas luces, el padre tiene razón cuando dice “mi hija me está tomando el pelo”. O al menos, me parece más que justificado que desconfíe de la respuesta de la hija.

¹⁰ La primera vez que encontré este ejemplo fue en un libro en portugués de Barry R. James, “Probabilidade: Um Curso em Nível Intermediário” (1981), donde se referían al mono como “o macaco”.

2. Margen de error total cuando la encuesta de intención de voto tiene un margen de error de +/- 3%. Tal como se dijo, la segunda etapa de “estimación el voto” no responde a reglas estadísticas ni probabilísticas, así que mal podemos hablar de varianzas, insesgamiento, error muestral, etc.

3. ¿De qué orden es la probabilidad de resultados tanto o más similares a los observados? En Estadística se usa frecuentemente el llamado valor-p (*p-value*) para someter a prueba una hipótesis. El valor-p trata de medir que tan improbables son los resultados obtenidos si aceptamos como cierta la hipótesis nula. En este caso la hipótesis nula sería: las encuestas de intención de voto son independientes (entre empresas y también entre meses para una misma empresa). El valor-p varía según sea el método utilizado para calcularlo, pero en todos los casos de la Sección 2 se observaron valores pequeños del mismo, tomando valores del orden de 6 en mil (Elliot considerando 3 meses), 2.5 en mil (Elices considerando cuatro meses) o 1.3 en diez mil (Elices considerando cuatro meses y dos de las semanas del balotaje). Sin embargo, por lo ya dicho, las consideraciones sobre el “margen de error total” me parece que no corresponden a conceptos probabilísticos y por lo tanto no afectan los cálculos.

4. El tipo de diseño no afecta los cálculos, siendo estos válidos en la medida que sea de aplicación el Teorema Central del Límite. En los diseños descritos por las empresas, y considerando tamaños muestrales del orden de 1000 casos, se cumplen las condiciones del Teorema Central del Límite para poblaciones finitas y podemos por lo tanto usar la distribución normal para calcular probabilidades¹¹.

11 Ver por ejemplo Francisco y Fuller (1991) “Quantile Estimation with a Complex Survey Design”, *The Annals of Statistics*, Vol. 19, No. 1 (Mar., 1991), pp. 454-469 o bien Isaki y Fuller (1982) “Survey Design Under the Regression Superpopulation Model”, *Journal of the American Statistical Association*, Vol. 77, No. 377 (Mar., 1982), pp. 89-96

5. Solicitar a las empresas la información necesaria para recalcular el margen de error estadístico. Es una interesante propuesta y con gusto lo haría, pero el error estadístico no cubriría la etapa de estimación de voto y por lo tanto el esfuerzo de aproximar las varianzas muestrales no sería relevante para este expediente.

6. (punto 5bis del escrito) Hay varios aspectos a considerar para responder si estadísticas de trigo, economía, tornillos o votos siguen las mismas reglas:
 - a. Naturaleza de la respuesta e instrumento de medida. Cuando nos referimos a toneladas de trigo o cantidad de tornillos el instrumento de medida es universalmente aceptado (balanza, contador de tornillos). Encuestas relativas a las ciencias sociales incluyen en general el uso de formularios elaborados por expertos en el área, formularios que pueden diferir de un investigador a otro y podrían ser criticados como no válidos para “medir” la opinión o la intención de una persona.
 - b. La posible retroalimentación que los resultados tienen sobre la población en estudio. El *feedback* que las sucesivas encuestas de intención de voto y de estimaciones de voto pueden tener sobre el comportamiento del electorado. Esta retroalimentación no ocurre en la producción de trigo o en el ingreso promedio de los hogares.

7. (punto 6 de escrito de Elices) Considero que los valores presentados por el Ing. Elices en la pág. 2 de su denuncia del 26 de noviembre de 2009, los cuales constituyen el origen del presente expediente, son resultado de “estimaciones de voto” y no de “encuestas de intención de voto”. En los términos del escrito de Elices los podríamos denominar “constructos producto de modelos de ciencias políticas” (foja 147 del expediente). En las referencias técnicas que las empresas hacen en sus sitios web:
 - a. CIFRA: En la referencia que hace CIFRA a su publicación “Intención de voto en octubre de 2009” de fecha 2 de octubre de 2009 (www.cifra.com.uy) dice: “Información sobre resultados de una encuesta nacional de CIFRA a 1006 entrevistados ...”.

- b. Equipos Mori: En el sitio web de Equipos, la referencia al artículo “ULTIMA PROYECCION DE EQUIPOS PREVIO A LAS ELECCIONES” aparece: “Los resultados y el análisis presentado en el presente informe fueron elaborados *sobre* la base del sistema regular de medición de Opinión Pública y Banco de Datos de Equipos MORI. En la última medición fueron entrevistadas 1200 personas ... entre los días 15 al 18 de octubre de 2009”.
 - c. En la exposición de Bottinelli en Radio El Espectador del jueves 22 de octubre (según transcripción en el sitio www.factum.edu.uy) bajo el título “¿Qué pasará el domingo? ¿Habrán balotaje? ¿El FA tendrá mayoría absoluta en el Parlamento?” se puede leer lo siguiente: “Podrá no gustarles nuestra actividad –incluso alguien ha llegado a decir “no existe la ciencia política”–, pero son encuestas científicas y de nivel universitario. No son concursos de acierto, no hay que jugarle a un número; son mediciones, aproximaciones que se registran. Y hay algo que siempre decimos y parece que es como oír llover: hay un margen de error estadístico, no es que Factum, Equipos o Cifras nos equivoquemos; las estadísticas tienen unas curvas que determinan el margen de error que puede dar una muestra, que en el caso de este tamaño es de +/- 3,2%”.
8. (sobre pregunta del punto 6 del escrito de Elices). En el punto anterior al menos dos de las empresas sostienen que los datos presentador están elaborados *sobre* resultados de encuestas. En el caso de Factum aparece como menos clara la etapa de “estimación de voto” y se hacen consideraciones del punto de vista estadístico.
9. (sobre pregunta del punto 7 del escrito) El Ing. Elices pregunta si una cuarta empresa podría alcanzar esos niveles de precisión. No sabría decirlo, en este punto no me considero como “experto actuante” sino como un ciudadano más. Desde esa perspectiva, mi opinión no técnica es que me inclino a decir que sí, de hecho hay otras empresas en el

mercado, aún cuando las tres seleccionadas en la denuncia son las que parecen tener mayor exposición.

6. Conclusiones

Se resumen los puntos que considero más importantes respecto a la denuncia del Ing. Francisco Elices.

1. Probabilidades pequeñas. En Estadística, si bajo los supuestos de una cierta hipótesis ocurre un suceso con probabilidad muy baja, esto nos hace dudar no de la ocurrencia del suceso, que es real, sino de la razonabilidad de las hipótesis bajo las cuales se calculó dicha probabilidad. En la Sección 2 ratifico los cálculos del Dr. Elliot y del Ing. Elices, cálculos que son la base de la denuncia. Bajo el supuesto de independencia entre encuestas los resultados observados son *demasiado* similares.
2. Encuestas de intención de voto y estimaciones de voto real. Pienso que donde aparece más clara esta distinción es en la pág. 6 del escrito del Ing Elices de fecha 8 de marzo de 2010 (el escrito figura a fojas 123 y sgtes. del expediente). Escribe Elices: “Las empresas encuestadoras realizan dos productos diferenciados: a) los resultados de la encuesta de intención de voto b) la estimación del voto.” La denuncia se basa en la independencia de las encuestas **pero** los resultados publicados **no** son los resultados de las encuestas, sino los resultados del proceso de “estimación del voto”. Como ya se explicó en la Sección 4, este proceso de estimación de voto no sigue reglas probabilísticas o estadísticas, por lo que no puede aplicársele el cálculo de varianzas, errores muestrales o probabilidades.
3. Falsa precisión. De acuerdo al Ing. Elices, algunas empresas incluyen en su publicidad referencias a la alta precisión de sus resultados. Creo que las empresas deberían hacer explícito (o más claramente explícito) que la precisión se refiere a la encuesta, pero que lo que se está publicando y comunicando a clientes y público en general es en realidad el resultado de la etapa de estimación de voto.

4. Encuestas independientes. Sin tener acceso a las encuestas o al menos al resultado de las mismas es imposible decir si se trata de encuestas independientes o no. Respecto a la segunda etapa, la de estimación de voto, estoy de acuerdo con lo expresado por las empresas, las tres empresas: las estimaciones de voto de cada empresa no son independientes entre sí, se usa la información que cada empresa va acumulando para mejorar sus estimaciones de voto en instancias futuras, la estimación de voto de cada mes está basada no sólo en la información de la encuesta de ese mes, sino en las estimaciones de los meses anteriores. Asimismo, está públicamente disponible lo que “se comenta”, lo que “se dice”, o lo que las otras empresas han publicado, si esto afecta o no a las estimaciones de voto de cada empresa me es imposible decirlo, no lo se.

En resumen, ¿las estimaciones de voto son independientes? No, no lo son dentro de cada empresa, y como dice Aguiar en su “explicación alternativa”, si hay “coordinación implícita, ex post” cuando proceden a construir sus estimaciones de voto, tampoco habría independencia entre empresas.

Apéndice 1. Programa de cálculo con la metodología usada por Dr. Michael Elliot.

```
# replica los calculos de Dr. Michael Elliot
# de acuerdo al email de 12 de octubre de 2009 (pag.9)

rm(list=ls())
set.seed(7551003)
R<-1000000
#R<-1e6 ##### para calcular un millón de veces

# resultados historicos
y<-c(44,45,44,
      45,46,45,
      44,44,45,
      45.5,46,46)/100
y

#####
#####
n<-9 # cuantos meses considero
(desobs<- sd(y[1:n]))
(medobs<-mean(y[1:n]))

x<-matrix(round(rnorm(n*R, mean=medobs, sd=(medobs*(1-
medobs)/1000)^0.5 ),
          digits=2), nrow=R)

system.time(desvio.esp<-apply(x,1,sd))

jpeg(file="Elliot 3 meses.jpg")
hist(desvio.esp,breaks=40, main="Desvíos muestrales esperados, 3
meses")
abline(v=desobs, lwd=5)
dev.off()
mean(desvio.esp< desobs)
mean(desvio.esp<=desobs)
desobs^2

#####
#####
n<-12 # cuantos meses considero
(desobs<- sd(y[1:n]))
(medobs<-mean(y[1:n]))

x<-matrix(round(rnorm(n*R, mean=medobs, sd=(medobs*(1-
medobs)/1000)^0.5 ),
          digits=2), nrow=R)

system.time(desvio.esp<-apply(x,1,sd))

jpeg(file="Elliot 4 meses.jpg")
hist(desvio.esp,breaks=40, main="Desvíos muestrales esperados, 4
meses")
```

```
abline(v=desobs, lwd=5)
dev.off()
mean(desvio.esp< desobs)
mean(desvio.esp<=desobs)
desobs^2
```


Apéndice 2. Programa de cálculo con la metodología usada por Ing. Francisco Elices.

```
# replica los calculos de Ing. Francisco Elices

rm(list=ls())
set.seed(7551320)
R<-1e6
# resultados historicos
y<-c(44,45,44,
     45,46,45,
     44,44,45,
     45.5,46,46)/100
y

#####
#####
n<-9 # cuantas encuestas considero
(desobs<- sd(y[1:n]))
(medobs<-mean(y[1:n]))

x<-matrix(round(rnorm(n*R, mean=medobs, sd=(medobs*(1-
medobs)/1000)^0.5 ),
          digits=2), nrow=R)

m<-n/3 # numero de meses considerados
terna<-matrix(x, ncol=3)

# considera éxito si diferencia <= 0.01
system.time(diff.esp<-apply(terna,1,max)-apply(terna,1,min))
diff.esp=round(diff.esp, digits=2)
exito<-0+(diff.esp<=0.01)
exit0<-0+(diff.esp< 0.01)

# calcula valor-p
mean(exito) # en cuantos meses pasa
ex<-apply(matrix(exito,ncol=m,byrow=R), 1, sum)
mean(ex==m) # en cuantos grupos de "m" meses pasa

mean(exit0) # en cuantos meses pasa
ex<-apply(matrix(exit0,ncol=m,byrow=R), 1, sum)
mean(ex==m) # en cuantos grupos de "m" meses pasa

jpeg(file="Elices cada mes.jpg")
hist(diff.esp,breaks=40, main="Desvío máximo esperado en cada
mes")
abline(v=desobs, lwd=5)
dev.off()

#####
#####
n<-12 # cuantas encuestas considero
(desobs<- sd(y[1:n]))
(medobs<-mean(y[1:n]))
```

```

x<-matrix(round(rnorm(n*R, mean=medobs, sd=(medobs*(1-
medobs)/1000)^0.5 ),
           digits=2), nrow=R)

m<-n/3 # numero de meses considerados
terna<-matrix(x, ncol=3)

# considera éxito si diferencia <= 0.01
system.time(diff.esp<-apply(terna,1,max)-apply(terna,1,min))
diff.esp=round(diff.esp, digits=2)
exito<-0+(diff.esp<=0.01)
exit0<-0+(diff.esp< 0.01)

# calcula valor-p
mean(exito) # en cuantos meses pasa
ex<-apply(matrix(exito,ncol=m,byrow=R), 1, sum)
mean(ex==m) # en cuantos grupos de "m" meses pasa

mean(exit0) # en cuantos meses pasa
ex<-apply(matrix(exit0,ncol=m,byrow=R), 1, sum)
mean(ex==m) # en cuantos grupos de "m" meses pasa

```

Apéndice 3. Artículo para determinar el número de iteraciones necesarias en una simulación.

Método Monte Carlo: determinación del número de iteraciones necesarias

Juan José Goyeneche

Instituto de Estadística
Facultad de Ciencias Económicas y de Administración
Universidad de la República

Noviembre de 2010

Resumen

El método de Monte Carlo se usa principalmente para estimar parámetros en modelos que involucran una o más variables aleatorias. En general estos modelos son muy complicados para resolver analíticamente.

Intuitivamente, cuantas más iteraciones se realicen, mejor será la precisión con que se pueden presentar los resultados. El objetivo de este artículo es presentar una forma de determinar R , el número de iteraciones necesarias.

Introducción

El método Monte Carlo consiste en simular el comportamiento de un modelo aleatorio a los efectos de aproximar la distribución en el muestreo de algunas estadísticas relacionadas con el mismo. La verdadera distribución en el muestreo de estas estadísticas es en general muy complicada de resolver analíticamente. Si bien el método está relacionado al uso de computadoras y la capacidad de éstas de generar números aleatorios, Daniel Peña menciona que el método ya fue usado por Student tomando 750 muestras de tamaño 4 de unas mediciones de dedos de 3000 delincuentes (Peña, (2001), pág. 40) a los efectos de estudiar la distribución t .

Número de Simulaciones y Precisión en los cálculos

Se simula el lanzamiento de dos dados, uno celeste y uno verde. Se consideran las variables aleatorias C_i y V_i , puntajes del dado celeste y del dado

verde en la i -ésima “tirada” respectivamente. Sean las variables aleatorias $A_i = I_{\{C_i \leq 5\}}$ y $B_i = I_{\{C_i \leq V_i\}}$. A lo largo de una serie de R tiradas los promedios $a_{(\cdot)} = R^{-1} \sum_{i=1}^{i=R} a_i$ y $b_{(\cdot)} = R^{-1} \sum_{i=1}^{i=R} b_i$ convergen en probabilidad a

$$\mu_A = E(A) = P(A = 1) = 5/6 \doteq 0,83333$$

y a

$$\mu_B = E(B) = P(B = 1) = (6 + 5 + 4 + 3 + 2 + 1)/36 = 21/36 \doteq 0,58333$$

respectivamente. En general los valores de $E(A)$ y $E(B)$ son desconocidos, así como los demás parámetros de la distribución de A y de B (y por eso es que hacemos simulación Monte Carlo).

La pregunta que pretendemos contestar es cual sería el valor de R necesario para aproximar los valores de μ_A y de μ_B con 3 o más cifras significativas con un 99% de confianza, o en general, con una precisión ε y una confianza de $1 - \alpha$.

Esto es, $P(|a_{(\cdot)} - \mu_A| > \varepsilon) \leq \alpha$ y $P(|b_{(\cdot)} - \mu_B| > \varepsilon) \leq \alpha$. Determinaremos primero el valor de R necesario para asegurar $P(|a_{(\cdot)} - \mu_A| > \varepsilon) \leq \alpha$.

Si R es suficientemente grande podemos aproximar las distribuciones de $a_{(\cdot)}$ con la distribución normal. El número de repeticiones necesarias surge entonces de hacer

$$\varepsilon / \sqrt{Var(a_{(\cdot)})} > Z_{1-\alpha/2},$$

donde $Var(a_{(\cdot)})$ es la varianza del promedio muestral de las a_i y $Z_{1-\alpha/2}$ es el cuantil $1 - \alpha/2$ de la distribución normal estandarizada. La $Var(a_{(\cdot)}) = \sigma_A^2/R$ puede aproximarse calculando $\hat{\sigma}_A^2 = (R - 1)^{-1} \sum_{i=1}^{i=R} (a_i - a_{(\cdot)})^2$, la varianza “muestral”¹. De las anteriores fórmulas surge que el número de repeticiones necesario para asegurar una precisión de ε con una confianza de $1 - \alpha$ es:

$$R > (Z_{1-\alpha/2})^2 \times \hat{\sigma}_A^2 / \varepsilon^2. \quad (1)$$

Con el “tamaño de muestra” determinado en 1 podemos asegurar la precisión y confianza requeridas para asegurar $P(|a_{(\cdot)} - \mu_A| > \varepsilon) \leq \alpha$. Podemos hacer lo mismo para determinar un valor de R que asegure la precisión y confianza requeridas para asegurar $P(|b_{(\cdot)} - \mu_B| > \varepsilon) \leq \alpha$ y luego tomar el supremo de ambos números.

Ejemplo.

Considere el siguiente código escrito en Matlab:

¹Para tener una primer aproximación a $\hat{\sigma}_A^2$ puede hacerse un número inicial de repeticiones (100, 500, ...).

```

r=1000
a=zeros(r,1);
b=zeros(r,1);
for i=1:r
    c(i)=unidrnd(6);
    v(i)=unidrnd(6);
    a(i)=(c(i)<=5);
    b(i)=(c(i)<=v(i));
end
abar=mean(a)
asig=var(a)
bbar=mean(b)
bsig=var(b)

```

El siguiente cuadro presenta los valores exactos y los valores estimados de los parámetros usando un número inicial de 1000 repeticiones.

Parámetro	Valor exacto	Valor estimado
μ_A	0,8333	0,8250
σ_A^2	0,1389	0,1445
μ_B	0,5833	0,5770
σ_B^2	0,2431	0,2443

Cuadro 1: *Valores aproximados y simulados*

Los valores de R necesarios para asegurar una precisión de 0.001 con una confianza del 99% para μ_A y μ_B son 921.591 y 1:612.951 respectivamente, por lo que en este caso deberíamos hacer al menos 1:612.951 iteraciones.

Otros parámetros.

En general la información producida por simulaciones Monte Carlo son promedios o funciones de promedios calculadas con R repeticiones del algoritmo. El siguiente cuadro presenta las fórmulas de las varianzas aproximadas de los estimadores por tipo de estimador. Todas las sumas que aparecen son para $i = 1, \dots, R$.

En el caso que los parámetros a presentar en los cuadros sean cuantiles u otras funciones que no puedan ser expresadas como funciones de promedios, las fórmulas de las varianzas del cuadro anterior no son aplicables.

Item reportado	Estimador	Varianza estimada
Media	$a_{(\cdot)} = R^{-1} \sum a_i$	$R^{-2} \sum (a_i - a_{(\cdot)})^2$
Raíz cuadrada de media	$(a_{(\cdot)})^{1/2}$	$4^{-1}(a_{(\cdot)})^{-1} R^{-2} \sum (a_i - a_{(\cdot)})^2$
Ratio de medias	$r = a_{(\cdot)}(b_{(\cdot)})^{-1}$	$(b_{(\cdot)})^{-2} R^{-2} \sum (a_i - rb_i)^2$
Raíz cuadrada de ratio	$r^{1/2}$	$4^{-1}r^{-1}(b_{(\cdot)})^{-2} R^{-2} \sum (a_i - rb_i)^2$

Cuadro 2: *Valores aproximados* de las varianzas de los parámetros simulados

Peña Sánchez de Rivera, Daniel (2001). “Deducción de distribuciones: el método de Montecarlo”, en Fundamentos de Estadística. Madrid: Alianza Editorial. ISBN 84-206-8696-4.