

PONENCIAS MESA 5

Javier Alagón

ESTADÍSTICA APLICADA PARA COPARMEX

Creo que se ha dicho demasiado con respecto a conteos rápidos y quisiera compartir con ustedes algunas de las reflexiones que me hago al final de los dos días ya de esta jornada. Vine aquí con muchísimas expectativas, algunas se han cumplido, otras no tanto, y quisiera ver hacia dónde vamos con esto que hemos aprendido a lo largo de estos dos días.

Decía Norbert Diner, premio Nobel de química, en su inglés con acento alemán, “*it’s difficult to predict in particular the future*”. Esto no parece ser para los conteos rápidos, realmente predecir con conteos rápidos es relativamente fácil sobre todo si se quita la restricción del tiempo, la restricción que tenemos que entregar a las nueve o diez de la noche cualquier cosa.

Antes que otra cosa les quería hacer un comentario con respecto al trabajo que todo el gremio realizó el 2 de julio, y creo que se ha soslayado. Creo que debemos estar realmente de jubilación: Muchísimas felicitaciones y lo quiero hacer en público, al trabajo hecho por el comité técnico, me pareció extraordinario lo que nos presentaron el día de hoy. Realmente no sabíamos que los mexicanos estábamos en tan buenas manos.

Ahora el conteo rápido. Ahí se presenta el objetivo; evidentemente es la estimación confiable, objetiva e independiente de la proporción de votos a cada uno de los partidos.

Nosotros realizamos un conteo rápido en 600 secciones electorales que correspondían a mil 315 casillas, reportamos creo que con 440 por ahí de las 22:15 horas. La selección de las secciones electorales fue de manera estratificada y, como ya se ha mencionado, por resultado de las votaciones en las elecciones presidenciales pasadas y por conglomerados dentro de cada estrato. Aquí sí nos quisimos tomar este riesgo, es una forma de llevar a cabo el muestreo que habíamos utilizado en dos ocasiones anteriores y que nos había funcionado de manera maravillosa, tanto en Jalisco como en Yucatán, en donde habíamos hecho los estratos con base en las votaciones anteriores.

La teoría dice que la estratificación funciona mejor, es decir, se disminuirá la varianza, conforme la variable de estratificación esté más relacionada con la variable que se está midiendo. Y aquí tenemos la situación ideal. No podíamos soslayar eso. Entonces, pues, para eso se realizó la estratificación, formamos siete estratos, creo que pueden llegar. En ocasiones previas habíamos tenido cinco estratos y en esta ocasión hicimos priístas muy priístas, panistas muy panistas y perredistas muy perredistas, competidas y secciones nuevas sin votación previa.

Los resultados del conteo rápido estuvieron listos a partir de las 10:15 con el 74 por ciento de las casillas computadas. Muy pocos de los que están aquí se enteraron de que hicimos un conteo rápido, fue el conteo rápido menos publi-

citado ¿Por qué?, porque además no tan solo este conteo rápido no se le dio tanta difusión, prácticamente se dio a las 20:00 horas con una encuesta de salida como lo hemos visto y, a partir de ahí, todo fueron muchísimas celebraciones y ya que si los números más o menos, la verdad, a la mayoría de los mexicanos no les importaba mucho.

Utilizamos 450 personas para reportar los resultados y nos instalamos en Coparmex en Insurgentes. Esto fue lo que reportamos a las 10:15 con 440.

Tenemos los resultados finales del IFE a la izquierda; a la derecha tenemos, en medio, los resultados que reportamos, nos quedamos 1.2 puntos porcentuales menos, 0.8 para el PAN, etc. El error de estimación en este momento estaba en 1.3 puntos porcentuales, esto es totalmente lo que sucedió: 1.2 por ciento, y tenemos un error de estimación de más menos 1.3; para el caso del PAN de más menos 1.1; para el caso del PRD más menos 1.2 puntos porcentuales.

Lo que habíamos tenido en experiencias anteriores, con un tamaño de muestra prácticamente bueno. Es más, incluso, la tercera parte en Jalisco y 200 secciones electorales en Yucatán, los márgenes de error estaban alrededor del 1.5 por ciento. Por eso pensábamos, junto con muchas simulaciones que hicimos para estimar los márgenes de error antes del día de la elección, todo tipo de pruebas, pensamos que iban a estar alrededor del 1.5 por ciento y se cumplió. Y los errores de estimación varían conforme los diferentes partidos.

Estas son las estimaciones puntuales, ya las vimos. Estas son las estimaciones por intervalo del conteo rápido, ya sin nulos, evidentemente, ahí le suman y le restan. Nos pasamos y aquí estoy usando decimales.

Este diseño metodológico, este diseño muestral como les comento, no son muchas experiencias, dos, pero había funcionado muy bien. En el caso de Yucatán esa sí fue una elección mucho más cerrada que ésta y en aquel caso, las estimaciones fueron más cercanas a lo que terminó el resultado final. Sin embargo, el margen de error fue muy parecido a lo que tuvimos aquí, fue el mismo nivel de precisión. En esta ocasión nos quedamos a un punto del PRI y del PAN, pero eso es simplemente cuestión de azar, no de precisión.

Tenemos, entonces, el marco muestral que tomamos, fue el listado de todas las secciones electorales, según el IFE, creo que esto me lo voy a brincar y pasamos a los siete estratos. Tenemos los tamaños de muestra de los diferentes estratos y el diseño es convencional, se los muestro y en la siguiente vemos que lo que hicimos fue utilizar las ponderaciones (WH) con base en 1994 y dado el carácter de elección presidencial decidimos utilizar 1994.

No creemos que sea un supuesto, a pesar de los comentarios de Edmundo Berumen y la discusión al respecto. Si bien es importante ver cómo se puede tratar el aspecto metodológico, hicimos muchas simulaciones con las muestras que sacamos, variando todo tipo de condiciones, variando tasas de participación dentro de las secciones electorales y sorprende la estabilidad de las estimaciones, es impresionante. Por eso creo yo que si trabajamos el problema, muy seguramente lo que nos va a dar es que a lo mejor no tiene gran efecto y que a lo mejor es un prurito académico y que vale la pena desde luego, da la impresión de que no va a tener mucho impacto.

Este es un diseño totalmente ortodoxo, totalmente documentado, las fórmulas están perfectamente desarrolladas en los libros y el chiste es echarlo a andar, ponerlo a funcionar, hacer que esto funcione y poder estimar los errores la noche de la elección para poder salir con cierto nivel de precisión.

Bien, hasta aquí del conteo rápido, fue la tercera ocasión, pero hasta ahí. Yo quisiera aprovechar estos últimos momentos para hacer unas cuantas reflexiones sobre lo que hemos vivido estos dos días.

Creo que no debíamos estar discutiendo sobre la ortodoxia de los métodos. Simplemente me dieron ayer una referencia (*flash back*) a la discusión del 1994, que si las muestras aleatorias, que si la muestra por conveniencia de no sé quién, que si las 25 ciudades de no sé quién más. Finalmente, estamos convencidos todos de qué debíamos utilizar y no debíamos salirnos del enfoque ortodoxo del muestreo.

Tenemos el asunto de la agenda pendiente, para mí esto es fundamental. ¿Cuándo se cruzaron las preferencias electorales? Nosotros hicimos una encuesta para un partido político de oposición en noviembre, créame que la diferencia entre el PRI y el PAN era enorme y entre el PAN y el PRD era también muy grande, eso fue en noviembre y no creo que haya cambiado mucho a enero. Entonces, eso sí me da la impresión de que en algún momento, y no porque yo haya hecho encuestas, de que por ahí de enero o febrero sí estaba el uno, dos, tres como estaba tradicionalmente, esa es mi impresión. Sin embargo, hemos visto exposiciones que, incluso, cuestionan esta posición, y la hipótesis, la premisa fundamental es, en algún momento, el PRI empezó a caer muy fuertemente y el PAN sube tanto que se cruzan las curvas.

¿Cuándo ocurrió eso?, pues sería lo de menos, ya ganó Fox y es lo de menos, pero no es así, porque para poder saber qué metodologías, qué filtros, qué preguntas, diferentes aspectos, necesitamos poder responder esta pregunta, si no la respondemos vamos a tener

la agenda pendiente y nos vemos en el 2006 y vamos a volver a tocar el mismo tema, y en el 2012 también.

Entonces, la pregunta fundamental es a qué se deben entonces las diferencias entre encuestas electorales. Me mete muchísimo ruido intelectual, pero la exposición que hizo ayer Eduardo Ragasol me mete muchísimo ruido en este sentido. Una agencia con el prestigio de Nielsen, con una contratación impecable, un medio, con la infraestructura de campo adecuada, con los muestristas adecuados, con los resultados que nos presentó, contra las otras posiciones que las otras encuestas que nos mostraban equis tendencia.

Eso desde luego que nos mete una interrogante, ¿qué es lo que no está funcionando?, sólo de esta manera si contestamos, vamos a poder proveer de información creíble a la sociedad mexicana, ya no digamos sólida, confiable, creíble, simplemente en términos de encuestas electorales.

Tenemos que dilucidar esto, son diferencias por diseños muestrales.

Sabemos que si empezaron en equis localidades y ahí se mantuvieron y esa pregunta no me la respondió más que Manolo, el punto de los que hicieron encuestas con rotación de muestras (*trackings*) espero que hayan hecho sus muestras independientes y no por abatir costos nos hayamos quedado o se hayan quedado en un punto geográfico similar. Tenemos que dilucidar estas diferencias por levantamiento de campo, diferencias por filtros, ya ha habido alguna pequeña discusión, lo que le falló a equis o zeta es la forma en que preguntó. Puede ser tan grave el poner totalmente probable, eso es lo que le estaba argumentando a Eduardo Ragasol, a poco eso fue lo único, realmente por eso no funcionó.

Bien, diferencias por posición de pregunta del voto, porque si no respondemos esto entonces el debate se va a ir hacia los indecisos, son los indecisos, estamos

en los indecisos o no, porque a lo mejor ni siquiera tenemos que llegar a los indecisos, a lo mejor las diferencias únicamente se deben a cuestiones de diseño, a cuestiones de campo, a cuestiones de cómo está la muestra y, en este sentido, ahora viene una propuesta.

La propuesta que yo quisiera hacer es, simplemente, juntar y analizar la información disponible; la pregunta es: ¿Estamos abiertos? Ayer Ulises Beltrán pasó un cuestionario, lo tiene Ana Cristina Covarrubias, y yo creo que habría que complementarlo para ver si lo vamos a responder. Pero vayamos un poco más allá, por qué no se sientan en la misma mesa Manolo y Gauss y Rolando Ocampo y ven cómo están las muestras geográficamente distribuidas, ven en qué distritos, en qué secciones y cómo fueron las votaciones en esas secciones, porque si ya desde ahí hay diferencias, caray, pues miren, ni nos peleemos con esa asignación de indecisos, ni siquiera nos vayamos por allá, ni nos peleemos. También por aquí se mencionaron sesgos favorables hacia el PRI, ahora ya parece que van a seguir hacia el PAN.

En la encuesta que nos presentó María de las Heras se había mencionado la parte de los sesgos hacia el PRI. No se ve, incluso, el sesgo hacia el PRI porque antes de la asignación de indecisos traía 11 ó 12 puntos porcentuales y ya después de la asignación de indecisos traemos 10 puntos porcentuales

Entonces, la pregunta es por qué desde 1994 había esta hipótesis de que si había un sesgo o no. Bueno, si hay, ¿lo podemos o no lo podemos quitar? Todo esto es una propuesta o la pregunta es ¿simplemente esperaremos a que la mano invisible del mercado, sabiamente, acomode las posiciones de los diferentes agentes que participan en las encuestas preelectorales? Y lo malo de todo esto es que esto va a ocurrir quizás en 18 ó 20 años, cuando hayamos ido aprendiendo poquito a poquito. ■