



Universidad Autónoma Metropolitana – Unidad Iztapalapa
19 de junio de 2018

Primer Foro Académico de Discusión Modelo de Estimación de Resultados Electorales Preliminares (MEREP)

<https://merep.izt.uam.mx>

Ricardo de la Peña



Luego de las votaciones del próximo 1 de julio de 2018, la ciudadanía contará de inmediato con diversas fuentes de información sobre los resultados de las elecciones, unas más exactas y oportunas que otras.

El Instituto Nacional Electoral tomará información para un conteo rápido de la elección presidencial a partir de los datos anotados en los llamados «cuadernillos de operaciones», por lo que por esta vía será factible contar con resultados de este ejercicio estadístico institucional en torno a las once de la noche del día de la elección. Este será sin duda el dato más sólido del que se dispondrá sobre el resultado de la elección para la Presidencia de la República.



Dados los tiempos esperados para que se culmine con los cómputos en las casillas el próximo 1 de julio, será sumamente difícil que se cuente con las estimaciones de otros conteos rápidos, no institucionales, al menos antes de la medianoche.

Pero se contará con el ya tradicional procedimiento de acceso público al acumulado de datos por casilla, conocido como Programa de Resultados Electorales Preliminares (PREP).



Empero, es sabido que este instrumento enfrenta un grave problema: dado que el arribo de los datos por casilla no es aleatorio, sino que responde a los tiempos de conclusión y trasmisión del cómputo de las votaciones en cada lugar, sus datos presentan sesgos derivados de diversos factores, como las diferencias en los husos horarios en las entidades, la condición urbana o no urbana de las casillas, otros elementos característicos de la población residente en cada región y complicaciones que se presenten en cada centro de votación. Pero ahora, dependerá además de cuántas elecciones se realicen en cada entidad, lo que variará entre sólo las tres federales o hasta seis votaciones.



Lo anterior hace que, antes de la medianoche del día de elecciones, no se tenga certeza sobre qué tan exactamente los datos acopiados en el PREP reproducirán el patrón de votación en la totalidad del país.

Sin embargo, los sesgos esperables en los datos del PREP y los posibles retrasos en la información sobre los resultados preliminares de la elección presidencial pueden ser resueltos mediante modelos alternativos de ajuste.



Es por ello que la Sociedad Mexicana de Estudios Electorales en coordinación con la Unidad Iztapalapa de la Universidad Autónoma Metropolitana y con la colaboración del Programa de Posgrado en Ciencias Políticas y Sociales de la Universidad Nacional Autónoma de México y del Capítulo México del Instituto Universitario de Investigación Ortega y Gasset, han decidido llevar a cabo y poner a disposición pública un modelo de estimación de los resultados electorales preliminares (MEREP).

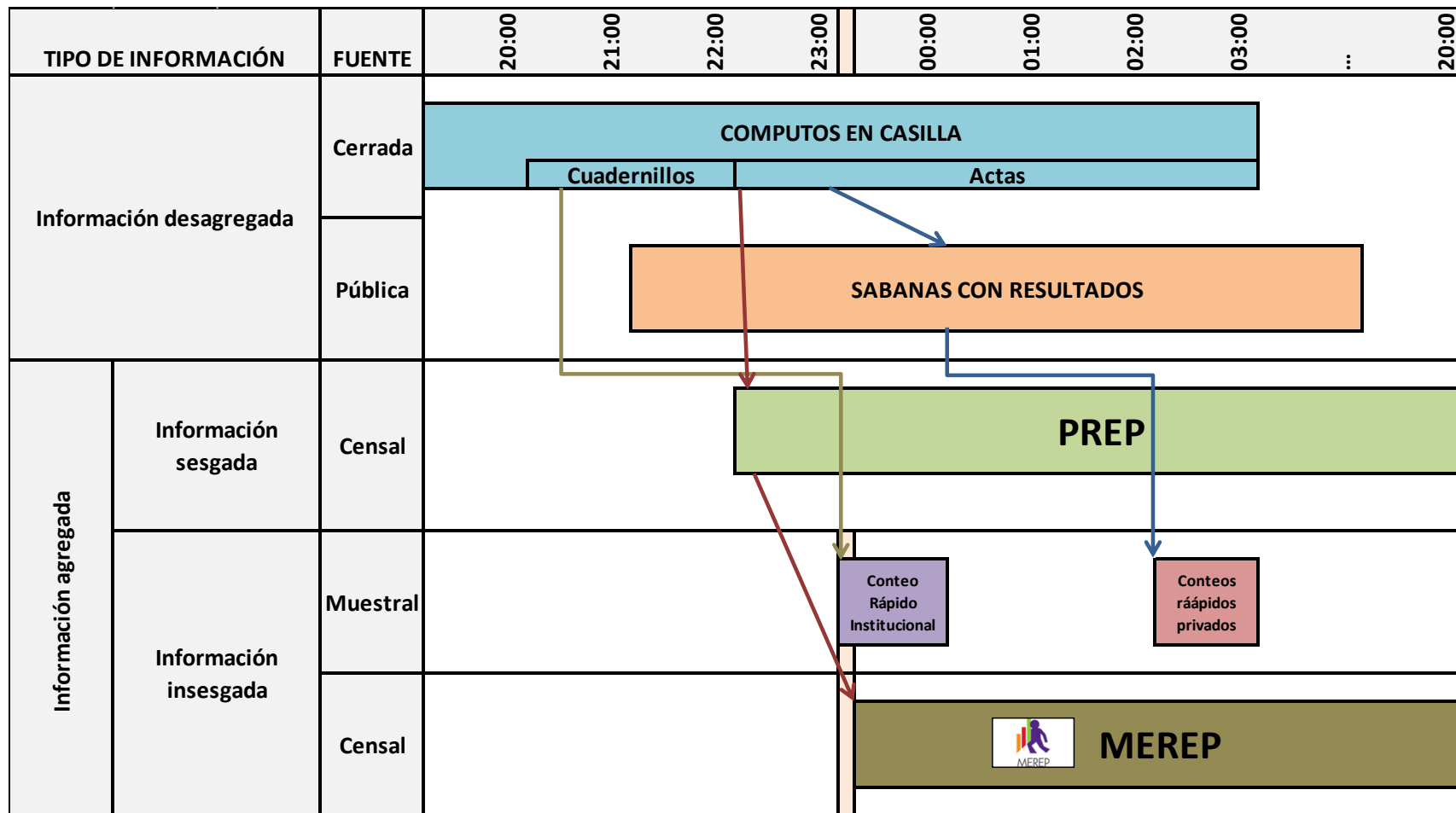


Este modelo asume una versión simple para los cálculos de los resultados, ya que no requiere de la extracción de muestras representativas de la base de datos acumulada por el PREP en un momento determinado, labor que supone procedimientos complejos y que demandan tiempo para su realización.

Esto se logra mediante la realización de una ponderación de la totalidad de los reportes acumulados en el PREP a un corte horario determinado, y no a una selección de tipo muestral de los mismos. Por ello, este ejercicio no es un conteo rápido ni una encuesta, sino una herramienta alternativa para contar con estimadores oportunos insesgados del resultado electoral.



Este modelo de estimación no constituye una fuente de información original de datos electorales, sino un derivado, cuya fuente son los datos que se van acumulando y son reportados en el PREP, por lo que el modelo se genera a partir de un programa que a su vez es alimentado por los datos computados en los centros de votación y que son captados y puestos a disposición de la ciudadanía a través del PREP. Por ende, no requiere disponer de un ejército de colaboradores recopilando datos afuera de casillas a lo largo y ancho del país.





Con este ejercicio se confía contar el mismo día de la elección, antes de medianoche, con datos de votación por contendiente con un error menor a un punto porcentual, dentro de niveles de confianza convencionales (95 por ciento), lo que permitirá disponer de tendencias confiables sobre el resultado de la elección presidencial de manera oportuna.

El modelo es enteramente transparente, puesto que su metodología estará a disposición pública y las estimaciones que sean difundidas en su oportunidad serán replicables y verificables, cumpliendo con los criterios fundamentales del hacer científico.



Este modelo de estimación será un instrumento adicional para disponer de **información agregada e insesgada** de resultados electorales preliminares, como el conteo rápido institucional, pero, a diferencia de éste, supone un ejercicio con pretensión **censal**, por lo que no demanda contar con una muestra definida para impedir o eliminar sesgos en la toma de datos debido a diferenciales en el arribo de datos por casilla por factores diversos.

La disposición de un modelo de estimación insesgada de los resultados electorales preliminares, como rutina derivada y dependiente del PREP, es una oportunidad para **acompañar el anuncio de resultados del conteo rápido institucional** con un ejercicio paralelo que constate dichos datos.

Se asume que, por los tiempos esperados para el arribo de la información al PREP, la apertura de las estimaciones producto de este modelo será posterior a la presentación pública de los resultados del conteo rápido institucional, que deberá hacer la autoridad administrativa electoral nacional en torno de las once de la noche del día de la elección.



A diferencia del conteo rápido institucional, este ejercicio de estimación es producto de la labor de diversas instituciones académicas por **contar de forma oportuna e independiente con estimaciones alternas** de los resultados de la próxima elección para Presidente de la República.

Es tan relevante esta elección que la disposición de datos y el cálculo de los resultados no debiera quedar confinada a una única fuente. En una democracia, es fundamental disponer de una diversidad de ejercicios que busquen aproximarse con exactitud al conocimiento temprano de las tendencias de la votación. Y quien mejor que la academia para sumarse a este esfuerzo.



¿Cómo se generaría esta estimación?

Se parte de distribuir las casillas en estratos, uno por distrito electoral federal y según condición urbana o no urbana de las casillas, y de estimar el volumen de electores en cada estrato, lo que es enteramente posible a partir de los datos del Listado Nominal de Electores. Luego, y conforme vayan arribando, se agruparían los datos reportados en el PREP según el estrato al que pertenezcan en cortes horarios regulares.



Después, simplemente se ajustaría el volumen de votos por contendiente en cada estrato, cuál si lo reportado en un corte horario correspondiera a la totalidad de la votación en dicho estrato, lo que se logra simplemente multiplicando los votos por contendiente contabilizados en un momento dado por el inverso de la proporción de electores cuyas casillas han sido reportadas en cada estrato hasta ese momento.

ESTIMADORES

Los estimadores para el cálculo de la proporción de voto para cada contendiente (\hat{p}_i) en un corte horario dado (t) resultarán de:

$$\hat{p}_{it} = \frac{\sum_{e=1}^m w_{iet}}{\sum_{i=1}^k \sum_{e=1}^m w_{iet}}$$

Donde

" \hat{p}_{it} " es la proporción estimada de votantes por un contendiente determinado en un corte dado;

" w_{iet} " son los valores de expansión de los casos en muestra en un estrato dado a un corte temporal específico;

" k " es el total de contendientes para los que se reportan datos en un corte dado; y

" m " es el total de estratos para los que se reportan datos a un corte determinado.

Y los valores de expansión de los casos en muestra vendrán dado por:

$$\text{Si } v_{iet} > 0 \rightarrow w_{iet} = v_{iet} \frac{r_{ie}}{r_{iet}}$$

$$\text{Si } v_{iet} = 0 \rightarrow w_{iet} = 0$$

Donde

" w_{iet} " es lo previamente indicado;

" v_{iet} " es la votación reportada en el PREP por un contendiente determinado (i) en un estrato dado (e) a un corte horario específico;

" r_{ie} " es el total de votantes registrados en el listado nominal en un estrato dado;

" r_{iet} " es el número de votantes registrados en el listado nominal en las casillas reportadas por el PREP en un estrato dado a un corte específico.

INTERVALOS

Los intervalos de confianza de cada estimador se calcularán con base en:

$$I_{1-\alpha} = (\hat{p}_{it} - e_{it}, \hat{p}_{it} + e_{it})$$

Donde:

" $I_{1-\alpha}$ " corresponde al intervalo de confianza al 100(1- α) por ciento, siendo en este caso " α " igual a 0.05;

" \hat{p}_{it} " es lo previamente indicado;

" e_{it} " es el error estadístico calculado para la estimación por un contendiente determinado en un momento dado.

Y los errores estadísticos de cada estimador (e_i) en un corte dado (e_{it}) serán calculados mediante la siguiente ecuación:

$$e_{it} = \sqrt{\frac{Z_{\alpha/2} N \hat{p}_{it} (1 - \hat{p}_{it})}{n(N - 1) + Z_{\alpha/2} \hat{p}_{it} (1 - \hat{p}_{it})}}$$

Donde:

" $Z_{\alpha/2}$ " corresponde al número de unidades de desviación estándar de la media a la que se estima el error muestral, que en este caso asume un valor de 1.96;

" N " es el total de casillas electorales establecidas en el territorio nacional;

" n " es el número de casillas electorales reportadas a un corte determinado;

y lo demás lo previamente indicado.



Teóricamente, este ajuste permitirá disponer de estimaciones que den cuenta de la **distribución más probable de los resultados preliminares** de la elección, que corregiría los sesgos implícitos en el PREP, que sólo tiende a aproximarse a los resultados totales de manera paulatina.

Empíricamente, puede establecerse que las proporciones de voto por candidato en las elecciones presidenciales previas estuvieron fuertemente relacionadas con el estrato en que se ubican las casillas y por tanto que hay diferencias significativas entre los grupos, que pueden entonces ser amortiguadas por medio de un ejercicio como el descrito.

VOTO EN LA ELECCIÓN PRESIDENCIAL 2006 POR ESTRATO (ANOVA)

		Suma de cuadrados	gl	Media cuadrática	F	Sig.
FCHr	Entre grupos	2218.240	536	4.139	406.756	.000
	Dentro de grupos	1187.824	116746	.010		
	Total	3406.064	117282			
RMPr	Entre grupos	905.134	536	1.689	233.706	.000
	Dentro de grupos	843.570	116746	.007		
	Total	1748.703	117282			
AMLOr	Entre grupos	2533.227	536	4.726	520.288	.000
	Dentro de grupos	1060.492	116746	.009		
	Total	3593.719	117282			
GRGr	Entre grupos	3.419	536	.006	43.510	.000
	Dentro de grupos	17.113	116746	.000		
	Total	20.531	117282			
PMCr	Entre grupos	15.543	536	.029	143.215	.000
	Dentro de grupos	23.639	116746	.000		
	Total	39.182	117282			

VOTO EN LA ELECCIÓN PRESIDENCIAL 2012 POR ESTRATO (ANOVA)

		Suma de cuadrados	gl	Media cuadrática	F	Sig.
JVMr	Entre grupos	1470.179	543	2.708	394.556	.000
	Dentro de grupos	954.603	139111	.007		
	Total	2424.783	139654			
EPNr	Entre grupos	912.354	543	1.680	201.408	.000
	Dentro de grupos	1165.791	139744	.008		
	Total	2078.145	140287			
AMLOr	Entre grupos	2293.008	543	4.223	536.066	.000
	Dentro de grupos	1100.830	139744	.008		
	Total	3393.838	140287			
GQTr	Entre grupos	15.238	543	.028	100.984	.000
	Dentro de grupos	38.423	138270	.000		
	Total	53.661	138813			



Es de explicitarse que este modelo asume el supuesto de que el mejor estimador del resultado en un estrato determinado y en un momento definido es el resultado observado en el conjunto de casillas de dicho estrato con datos reportados hasta ese momento, toda vez que la dispersión en las proporciones de votación por candidato entre casillas dentro de un mismo estrato tenderán a ser menores que cualquier dispersión posible para subconjuntos de casillas que pudieran definirse antes de conocerse los resultados.



Esto, debido a que la decisión sobre las particiones del universo de casillas en estratos considera los criterios objetivos y disponibles que pueden dar cuenta de mejor manera de las diferencias en las proporciones de votación entre casillas dentro del territorio nacional, sin considerar supuestos exógenos a los datos mismos que se acopian y a los resultados que se pretenden estimar: su división en distritos y la condición urbana o no urbana de las unidades dentro de cada uno de ellos.

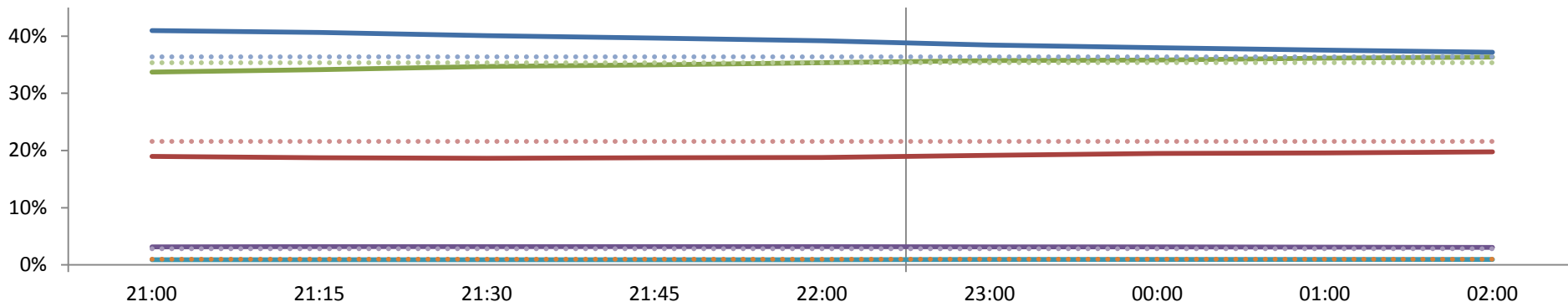


De hecho, se puede considerar que la estrategia propuesta permitirá compensar los eventuales sesgos derivados no sólo de las diferencias en husos horarios, sino sobre todo las esperables dilaciones en el cierre de cómputos en entidades donde se realicen elecciones federales y locales de manera concurrente.

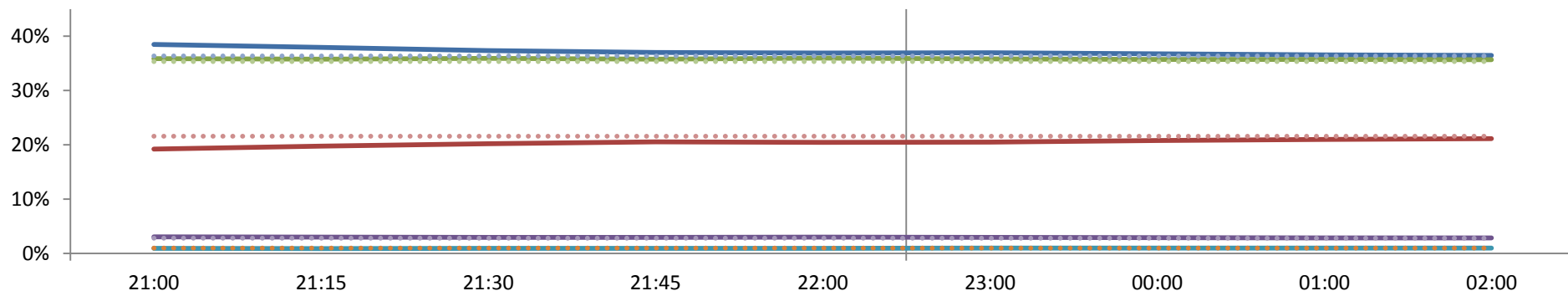
Así, con este modelo se contará desde el arranque con estimaciones próximas al resultado preliminar de la elección, aunque desde luego la concordancia irá mejorando con el tiempo.



Programa de Resultados Electorales Preliminares Elección para Presidente de la República 2006 según horario

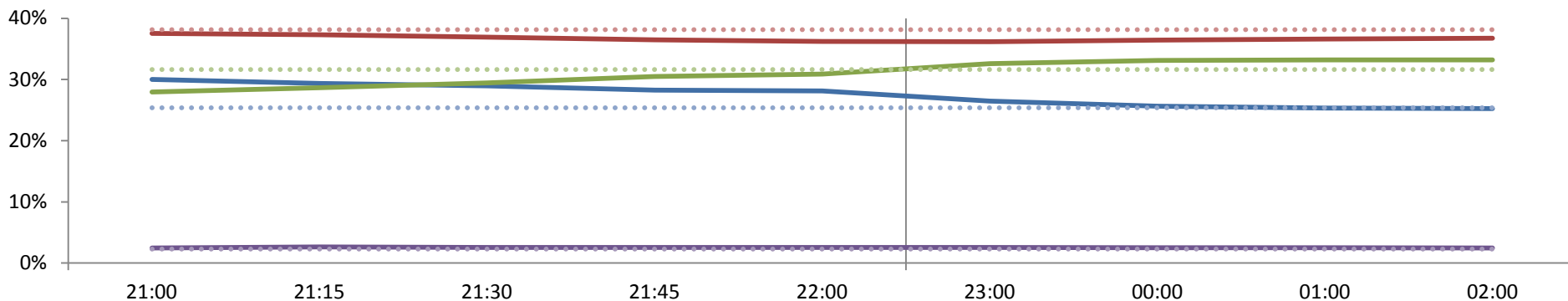


Modelo de Estimación del Resultado Preliminar de la Elección para Presidente de la República 2006 según horario

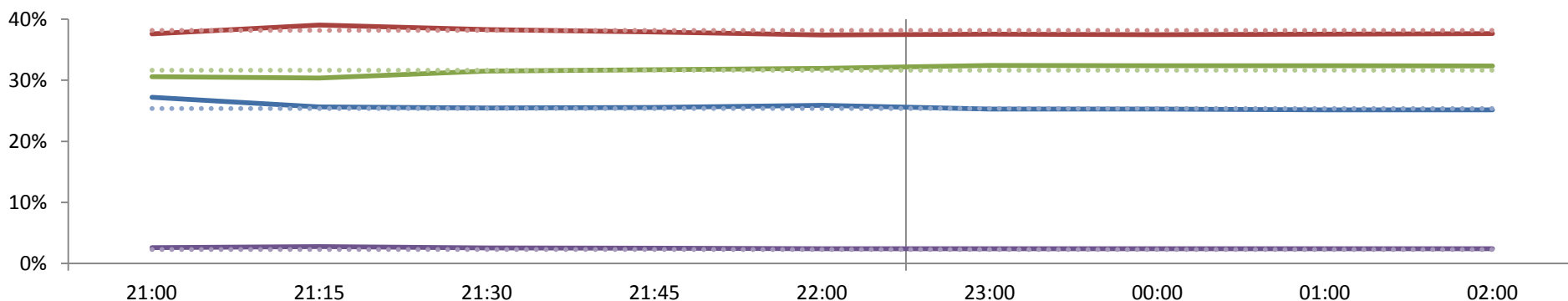


— FCH — RMP — AMLO — GRG — PMC

Programa de Resultados Electorales Preliminares Elección para Presidente de la República 2012 según horario



Modelo de Estimación del Resultado Preliminar de la Elección para Presidente de la República 2012 según horario



JVM EPN AMLO GQT



Para estimar un horario en el que se podrá disponer de datos consolidados y estables para la eventual apertura pública de las estimaciones producto del modelo, es menester considerar avances establecidos por las autoridades para la recopilación rápida de los resultados de las actas de escrutinio y cómputo y su inclusión en el PREP, pero también retrasos que pudieran derivarse de la integración en una casilla única donde hay elecciones federales y locales concurrentes.



Esto lleva a considerar como expectativa racional que el arribo abundante del flujo de información requerido para una estimación estable y precisa sea **después de las once de la noche y antes de las doce de la noche del 1 de julio de 2018.**

Por estimación precisa se entendería que los intervalos para cada contendiente no sean mayores a un punto porcentual por encima o por debajo de la estimación puntual. El reporte dará en todo caso cuenta de las **estimaciones puntuales y de los intervalos** (al convencional 95% de confianza) calculados en cada corte, que estarán programados cada quince minutos.



Modelo Responsable Participantes



Modelo de **Estimación del Resultado Preliminar** de la Elección para Presidente de la República 2018 (MEREPE)

Los datos que aquí se presentan son resultado de un ejercicio de simulación basado en el PREP y no necesariamente corresponden con las proporciones reportadas por este programa. Los resultados de este ejercicio no tienen carácter oficial bajo ninguna circunstancia. Los resultados oficiales de las elecciones federales son exclusivamente aquellos que dé a conocer el Instituto Nacional Electoral y, en su caso, el Tribunal Electoral del Poder Judicial de la Federación.

Última actualización:

Corte a las 20:00 horas (UTC-5),
 lunes 2 de julio de 2012

Actas capturadas:

141,935 (98.95%)

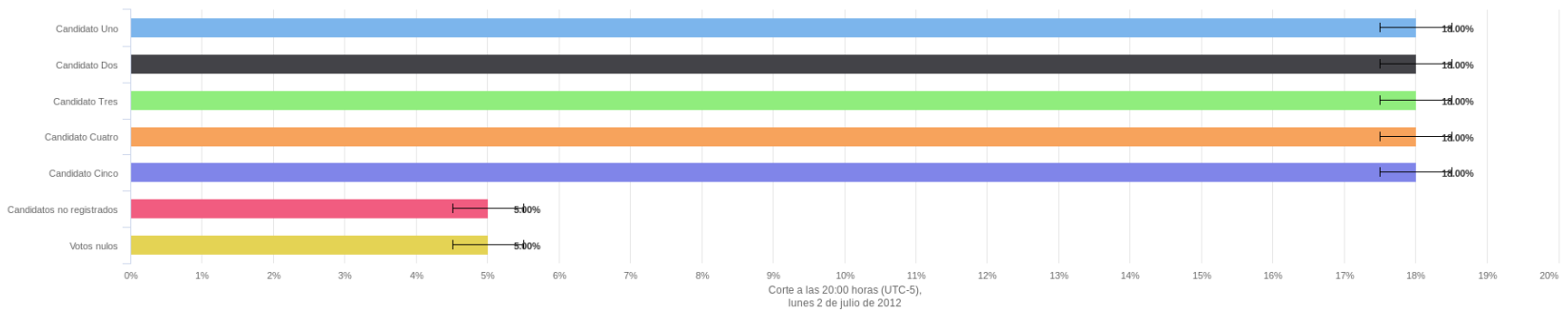
Lista Nominal:

98,418,243

Participación Ciudadana:

63.14%

Estimación del Resultado Preliminar para cada candidato a nivel nacional en porcentaje





Al momento marcado para la apertura del sistema es posible que se cuente con reportes de al menos **3% de las casillas** del país, siendo esta cota la mínima deseada para abrir los datos al público, garantizando así confiabilidad en las estimaciones que se difundan.

Una vez abierto el sistema, al igual que el PREP y a diferencia de los ejercicios de conteo, la estimación se dejaría «correr», por lo que se **actualizarán los datos en cortes programados y automatizados** hasta el cierre del propio PREP.



Al cierre, los datos del modelo apenas diferirían de los del PREP, pero las diferencias que hubiera, en todo caso, tenderán a aproximar las estimaciones del modelo con los resultados de los cómputos definitivos, al compensarse las casillas cuyos resultados estén todavía ausentes al cierre del PREP.

Esto es una manera sencilla, diáfana y sin supuestos riesgosos sobre el patrón de votación de los electores, que permite acercarse al resultado real de la elección de forma temprana. Esperemos que ello se constate el próximo domingo 1 de julio.



Modelo de Estimación del Resultado Preliminar de la Elección para Presidente de la República 2018

<https://merep.izt.uam.mx>