CAPÍTULO CINCO

Principios de Estadística y el Conteo Rápido

La metodología del conteo rápido aplica principios de estadística a un problema sumamente práctico: verificar un resultado electoral.[33] Este capítulo esboza dichos principios estadísticos y describe cómo operan juntos. La forma más breve de presentar esta información es utilizando el lenguaje de las matemáticas, y éste hasta cierto punto es inevitable. Sin embargo, el objetivo de este capítulo es presentar los conceptos básicos en forma nada técnica para que la lógica de la metodología del conteo rápido sea accesible al público en general.

La primera parte de este capítulo presenta las bases de la metodología del conteo rápido. Comienza examinando la solidez de los datos del conteo rápido y conceptos centrales tales como muestra y población. Luego se pasa a explicar principios estadísticos tales como la ley de los grandes números y el teorema del límite central. La segunda parte del capítulo, más técnica, presenta el proceso de construcción de una muestra. Esta sección esboza la medición de la tendencia central y la dispersión, y luego discute las estrategias usuales para calcular y extraer muestras. También se refiere a cuestiones prácticas tales como los factores de corrección, diseñados para manejar los problemas singulares que surgen durante la aplicación de principios estadísticos a situaciones de conteo rápido.

PRINCIPIOS FUNDAMENTALES DE LA ESTADÍSTICA

Los principios de la estadística impulsan la metodología para la recolección y el análisis de los datos de un conteo rápido. Esta metodología se basa en principios científicos ampliamente aceptados. Al igual que la gravedad, estos principios no son una simple cuestión de opinión ni tampoco están abiertos a interpretaciones partidistas; son demostrables y universalmente aceptados.

Los principios de la estadística impulsan la metodología para la recolección y el análisis de los datos de un conteo rápido.

[33] Este capítulo se concentra en los principios estadísticos involucrados en la extracción de una muestra aleatoria de las mesas de votación, en las cuales se reúnen y analizan los datos para proyectar o verificar los resultados de la elección. Sin embargo, la metodología del conteo rápido ha evolucionado y algunos de los mismos principios estadísticos ahora impulsan la observación cualitativa de una elección. El Capítulo 6, El Componente Cualitativo del Conteo Rápido, describe cómo la información de los procesos de votación y conteo puede recogerse de los mismos observadores y las mismas mesas de votación utilizados para recuperar datos sobre el proceso de sufragio. Estos hallazgos pueden generalizarse con confianza en relación a la calidad de los procesos de votación y conteo de todo el país.

La metodología del conteo rápido permite a un grupo mostrar por qué razón los procesos del día de la votación pueden considerarse justos, o en qué medida fueron injustos.

Es precisamente debido a que estos principios tienen base científica que los organizadores de un conteo rápido pueden señalar con autoridad los resultados electorales. Una cosa es sostener que una elección fue justa o injusta. Pero la metodología del conteo rápido permite a un grupo mostrar por qué razón los procesos del día de la votación pueden considerarse justos, o en qué medida fueron injustos.

Confiabilidad y Validez

Las afirmaciones que se hagan sobre los procesos del día de la votación son tan sólidas como los datos en los cuales se basan. Por lo tanto, es importante tomar medidas bastante deliberadas para asegurar que la información reunida cumpla con ciertos requisitos. Uno de ellos es que los mismos datos del conteo deben ser "sólidos". Esto significa que los datos deben ser tanto confiables como válidos.

Se considera que los datos son confiables cuando los observadores independientes que evalúan el mismo evento (el conteo de los votos) usando el mismo instrumento (el formulario del observador), expresan exactamente el mismo juicio. Un ejemplo simple muestra esto:

Tres personas distintas (A, B y C) miden repetidas veces la altura de una cuarta persona (Z) el mismo día. La medida de la altura de dicha persona será considerada confiable si los tres observadores (A, B y C), usando el mismo instrumento de medición (una cinta de medir estándar), obtienen exactamente el mismo resultado al medir a Z.

Es importante tomar medidas bastante deliberadas para asegurar que los datos reunidos cumplan con ciertos requisitos. Uno de ellos es que los mismos datos del conteo debe ser "sólidos".

El mismo principio se aplica a la recolección de datos para el conteo rápido; es esencial que tanto los indicadores como las medidas sean confiables. La información producida por los observadores no deben cambiar por los indicadores pobres, instrumentos de medición inadecuados (una cinta de medir elástica) o procedimientos malos, y los resultados tampoco deberían variar dependiendo de quién toma la medida. Los resultados confiables varían únicamente cuando hay cambios genuinos en el fenómeno que se está midiendo. Por lo tanto, datos confiables son aquellos que se pueden verificar independientemente.

Los datos del conteo rápido deben asimismo ser válidos. La validez se refiere a qué tan bien cualquier indicador a usarse encaja realmente con el concepto que se está midiendo. Se considera válida una medida si el indicador utilizado para su medición corresponde exacta e íntegramente al ámbito y el contenido del objeto que está siendo medido. Para ejemplificar esto podemos ampliar el ejemplo anterior:

A tres observadores adicionales (D, E y F) se les pide que reporten el tamaño de la misma persona, Z. D y E tal vez digan que Z, que mide 1.80 m, es grande, en tanto que F podría decir que es de talla mediana. El problema radica en que el indicador de la talla es ambiguo y está abierto a distintas interpretaciones; para algunas personas podría significar algo más que la altura; por lo tanto, el tamaño carece de validez. D podría considerar grande a Z porque éste es mucho más alto que él mismo. E podría pensar que Z es grande porque es más robusto en comparación a él. F tal vez informe que Z tiene tamaño mediano porque tienen la misma talla y altura, y F se considera una persona de tamaño mediano. En realidad, la ambigüedad de la noción de tamaño es un problema y constituye una amenaza para la confiabilidad y la validez.

Es por estas razones que las encuestas a boca de urna y las encuestas de opinión deben interpretarse con extrema cautela. Ellas a menudo producen estimados de los resultados reales del sufragio el día de las elecciones que no son confiables. Esto se debe a que las encuestas a boca de urna miden los recuerdos y las de opinión la intención de voto de los ciudadanos. Por razones sumamente comprensibles, las personas se ven tentadas a tergiversar ya sea cómo votaron, o bien cómo piensan hacerlo. En cambio, el conteo rápido mide el comportamiento, no los recuerdos o la intención anunciada. El conteo rápido mide cómo las personas realmente votaron, no cómo podrían haber reportado su voto a un extraño.[34]

PREGUNTAS FRECUENTES

¿Cómo pueden asegurarse los líderes del conteo rápido de que sus datos son confiables y válidos?

El primer paso es asegurarse de que todos los líderes y el personal entiendan estos conceptos. Todos los que están involucrados deberían ser conscientes de que los resultados del conteo rápido se verán comprometidos si los datos no son confiables (verificables independientemente) y válidos (miden lo que se busca medir). El segundo paso es describir las implicaciones. Por ejemplo, debe probarse la validez de las preguntas en los formularios de observación, y estos formularios deben incluir categorías de respuestas que permiten que los datos sean reportados de manera confiable. Los programas de capacitación son igualmente cruciales. Deben ser diseñados de manera que aseguren que todos los voluntarios entiendan aobservadores medirán el mismo evento utilizando el mismo formulario de la misma manera.[35]

[34] El conteo rápido mide también aspectos cualitativos del proceso de votación y conteo y, como se discute en el Capítulo 6, *El Componente Cualitativo del Conteo Rápido*, hay que tener mucho cuidado en el diseño de las preguntas para medir los indicadores cualitativos.

[35] Estos temas se consideran con mayor detenimiento a lo largo de todo el manual. El Capítulo 4, *Construyendo la Red de Voluntarios*, examina el diseño de los formularios y manuales de observación, y el reclutamiento y la capacitación de voluntarios. El Capítulo 6, *El Componente Cualitativo del Conteo Rápido*, presenta aún más recomendaciones para el diseño de los formularios de observación del conteo rápido.

Las muestras del conteo rápido brindan una base confiable con la cual formular estimados precisos de la población total.

El conteo rápido parte del supuesto de que los datos del conteo de los votos son confiables y válidos.

La Muestra

La solidez de los datos del conteo rápido también depende de cómo se construye la muestra; ésta determina qué votos se usan como base para estimar el resultado de una elección. La idea fundamental de una muestra aparece de muchas formas distintas en la vida cotidiana. Por ejemplo, los químicos rutinariamente toman una "muestra" de un compuesto y la analizan para formular enunciados exactos sobre las propiedades químicas de todo el compuesto. Los médicos toman muestras de sangre de los pacientes para establecer si la composición de su sangre está causando una enfermedad. Afortunadamente no necesitan extraer toda la sangre del cuerpo de sus pacientes para saber exactamente qué contiene. Tal enfoque no es práctico y es innecesario, puesto que una sola muestra de sangre revela todo lo que el médico necesita saber del contenido de toda la provisión de sangre del paciente.

La muestra del conteo rápido se basa exactamente en los mismos principios. Un grupo de observación podría considerar pedir a los voluntarios que observen toda mesa de votación del país y que reporten todos los resultados. Esta estrategia requeriría una gran cantidad de recursos y es innecesaria. Al igual que el químico y el médico, los grupos de observación pueden obtener todo lo que necesitan saber sobre toda la población votante empleando una muestra cuidadosamente diseñada. El método es más rápido, económico y práctico.

Las muestras del conteo rápido brindan una base confiable con la cual formular estimados precisos de la población total, dado que una muestra es un subconjunto particular de la misma, un subconjunto que revela características de la población. Aún así, el diseño de una muestra requiere ciertas selecciones, y éstas pueden tener un profundo impacto sobre la precisión de los datos asi como sobre los tipos posibles de análisis de la información.

La Población

Técnicamente, una población se refiere a todos los casos individuales relevantes existentes dentro de cierto límite. A los estadísticos a menudo no les interesa contar individuos. Al conteo rápido no le interesan todas las personas que viven dentro de las fronteras de un país particular. Solamente le preocupa la población relevante: toda persona que puede votar.

La población relevante del conteo rápido excluye a todas las personas que no están habilitadas para votar, sea por la razón legal que fuere. Las leyes electorales de la mayoría de los países tienen reglas muy claras sobre, por ejemplo, la edad en que se puede votar. Por lo general, las personas muy jóvenes no están habilitadas para votar, aunque el límite de edad preciso varía de un país a otro. Del mismo modo, la mayoría de los países tienen requisitos de ciudadanía que únicamente permiten votar en las elecciones nacionales a los ciudadanos.[36]

[36] Hay que señalar que la naturaleza democrática de una elección puede ser negada por exclusiones impropias y discriminatorias del derecho a votar y / o por la manipulación del padrón oficial de votantes. Estos puntos no son tratados por el conteo rápido pero deben ser cubiertos por otras actividades oficiales de observación de elecciones. Véase, por ejemplo, Desarrollando la Confianza en el Proceso de Inscripción de los Votantes: Una Guía del NDI para Partidos Políticos y Organizaciones Cívicas, 2001.

Pasando de una Muestra a una Población

El conteo rápido parte del supuesto de que los datos del conteo de los votos son confiables y válidos. En otras palabras, el conteo rápido asume que el conteo oficial efectuado en las mesas de votación —la información reunida por los observadores de todos y cada uno de los puntos de muestra— es información sólida. De hecho, los grupos de observadores logran verificar dicho supuesto emprendiendo una observación cualitativa sistemática de los procesos de votación y conteo en las mesas de votación.[37]

Si una observación cualitativa sistemática de los procedimientos del día de las elecciones establece que los datos del conteo de votos son confiables y válidos, y si se siguen los principios fundamentales de estadística, entonces pueden realizarse estimados precisos de la distribución de los votos por todo el país en base a una muestra adecuadamente construida. Gracias a la teoría de las probabilidades es posible efectuar estimados sumamente exactos del comportamiento de una población (cómo votó ésta) a partir de una muestra (de los resultados en mesas de votación escogidas).

Probabilidades: La Ley de los Grandes Números y el Teorema del Límite Central

La probabilidad se refiere a la posibilidad de que un evento o un resultado se produzca. Es posible estimar la probabilidad de eventos futuros y desconocidos —que Brasil ganará la Copa del Mundo, o que lloverá hoy. Nadie sabe lo que sucederá por anticipado, pero es posible hacer una suposición informada en base al rendimiento del equipo en otros eventos, o las condiciones meteorológicas afuera. Es igualmente posible hacer predicciones sobre probabilidades en base a la posibilidad conocida de que algo sucederá. Considérese el clásico ejemplo estadístico de arrojar una moneda, una que no esté cargada:

Una moneda se tira al aire 100 veces. Con una moneda buena, las posibilidades son que el resultado sea cara 50 veces y sello 50 veces, o algo muy cerca de ello. Supóngase ahora que la misma regla se probara con sólo unas cuantas lanzadas de la misma moneda. Tirar esa misma moneda 12 veces al aire podría dar resultados que no son exactamente parejos. El resultado podría ser 9 caras y 3 sellos. De hecho, en circunstancias excepcionales, es posible que en las 12 veces que se tire la moneda, resulte cara en todas ellas. En realidad, la probabilidad de que se de un resultado tan inusual puede calcularse con bastante precisión. La probabilidad de que salgan doce caras una tras otra resulta ser de una de dos a la duodécima potencia (1/2)12, o una de 4,096 o 0.024 por ciento. Esto es, la posibilidad de sacar doce caras (o sellos) en fila es una de cuatro mil noventa y seis. La teoría de las probabilidades indica que la distribución de la aparición de caras y sellos se igualará a largo plazo.

Un aspecto de la teoría de las probabilidades que opera en el ejemplo anterior del lanzamiento de la moneda es la ley de los grandes números. Este principio

[37] El Capítulo 6, El Componente Cualitativo del Conteo Rápido, detalla los procedimientos seguidos para evaluar sistemáticamente la calidad de los procesos de votación y conteo.

estadístico sostiene que cuantas más veces se arroje una moneda buena al aire, tanto más probable es que la distribución general de los resultados totales (observaciones) se adecúe a un patrón totalmente conocido y predecible. La implicación práctica queda clara: cuantos más datos tengamos, tanto más seguros estaremos de poder predecir resultados con precisión.

Esta ley estadística cuenta con una base firme en la matemática, pero la lección en términos laicos es que hay seguridad en los números. Un segundo ejemplo ilustra un punto relacionado que es importante para comprender los fundamentos de la metodología del conteo rápido.

Considérese una clase de 500 alumnos que toman el mismo curso universitario. La mayoría de ellos tendrán B y C, aunque unos cuantos obtendrán A y unos pocos D o hasta F. Casi con toda seguridad, esa misma distribución de calificaciones no se repetirá exactamente si el mismo curso cuenta con una clase de 10 alumnos o menos. Aún más importante es que las calificaciones de alumnos excepcionalmente buenos o malos tendrá un impacto sumamente distinto en la calificación promedio de toda la clase. En una clase pequeña, estas calificaciones "periféricas" tendrán un gran impacto sobre la distribución general y el promedio de la clase; ellas sesgarán los resultados de la curva de calificaciones. Pero en una clase más grande, el impacto de toda calificación excepcional tendrá un impacto bastante menor sobre la calificación promedio de toda la clase.

Cuanto más grande sea el número de observaciones con que contemos, tanto más probable será que podamos hacer predicciones estadísticas confiables sobre las características de la población.

La implicación práctica del ejemplo de la distribución de calificaciones es simple: a medida que la cantidad de datos (la cantidad de puntos de observación) se incrementa, disminuye el impacto que cualquier punto de observación individual tiene sobre el resultado total.

Un segundo principio estadístico que es vital para la metodología del conteo rápido se conoce como el teorema del límite central. Este axioma sostiene que cuanto más grande sea el número de observaciones (puntos de muestreo), tanto más probable será que la distribución de los puntos de datos tienda a seguir un patrón conocido. Una clase de 500 estudiantes de física en Brasil arrojará la misma distribución de calificaciones que una clase de 300 alumnos de literatura en Francia, aun cuando las calificaciones mismas puedan ser distintas. En ambos casos, la mayoría de los puntos de datos tenderán a concentrarse alrededor de la calificación promedio.

Estos dos axiomas estadísticos — la ley de los grandes números y el teorema del límite central — trabajan conjuntamente. Juntos indican que:

1. cuanto más grande sea el monto de observaciones (puntos de muestreo), tanto menos probable será que un resultado excepcional afecte al promedio (ley de los grandes números); y
2. cuanto más grande sea el número de observaciones, tanto más probable será que el conjunto de datos como un todo produzca una distribución de casos que corresponda a una curva normal (teorema del límite central).

De estas reglas estadísticas sale un principio general que tiene poderosas implicaciones para los conteos rápidos: cuanto más grande sea el número de observaciones con el que contemos, tanto más probable será que podamos hacer predicciones estadísticas confiables sobre las características de la población. Sin embargo, es absolutamente crucial entender que para que estos dos principios estadísticos se cumplan, la selección de los casos en la muestra debe ser aleatoria.

Para que la muestra resultante sea representativa de la población total, los puntos de la misma deben elegirse al azar y sólo al azar.

Aleatoriedad

Una muestra puede ser pensada no sólo como un subconjunto de una población, sino también como una réplica en miniatura de la población de la cual fue extraída. La población de todo país puede ser considerada única en ciertos aspectos. No hay dos países que sean iguales cuando se trata de ver cómo características tales como la lengua, la religión, el género, la edad y la educación se distribuyen en toda la población. Ya sea que una persona tenga carro, viva en una ciudad y no en un pueblo, tenga trabajo o tenga un perro, todo ello contribuye a la singularidad de la experiencia personal. Es imposible preparar una lista definitiva y exhaustiva de todas y cada una de las características singulares que nos distinguen como individuos, y mucho menos para toda una población; simplemente hay demasiadas combinaciones posibles de factores qué documentar. Afortunadamente, el conteo rápido no requiere eso. A éste no le interesan todas las cosas que hacen que las personas sean distintas. Al conteo rápido únicamente le interesan factores que tienen un impacto demostrable sobre la distribución de votos dentro de la población de votantes.

Para que la muestra resultante sea representativa de la población total, los puntos de muestra de la población relevante deben elegirse al azar, y sólo al azar. En la práctica, la aleatoriedad significa que la probabilidad de que un punto cualquiera de la muestra de la población sea elegido, es exactamente la misma que la probabilidad de que otro punto de la muestra lo sea. Y por las razones que ya hemos señalado, la ley de los grandes números y el teorema del límite central indican que cuanto más grande sea la muestra trazada, tanto más grande será la precisión con la cual dicha muestra representará las características de la población.

Cuanto mayor sea la heterogeneidad de la población votante, tanto más grande tendrá que ser la muestra a fin de arrojar un estimado preciso del comportamiento electoral.

Homogeneidad y Heterogeneidad

Las muestras confiables no requieren una inmensa cantidad de información detallada sobre las características sociales de la población total. Sin embargo, es esencial saber si la población que nos interesa es relativamente diversa (heterogénea) o no (homogénea). La evaluación de la heterogeneidad y la homogeneidad tiene un impacto significativo sobre cómo muestrear una población de manera confiable.

Hay varias formas de examinar el nivel de heterogeneidad, o diversidad, de toda población. La composición étnica, la religión y la lengua pueden afectar la heterogeneidad. Sin embargo, la preocupación principal del conteo rápido no es simplemente el nivel de heterogeneidad étnica o religiosa de una población. La pregunta crucial para el conteo rápido es si dicha heterogeneidad tiene un impacto significativo sobre el comportamiento electoral. Se considera que una población es relativamente homogénea si un candidato es preferido por el 80% de la población, sin importar la diversidad religiosa, lingüística o étnica de la misma. De igual modo, se considera que una población es relativamente heterogénea si la carrera electoral es estrecha, estando los votos divididos entre dos o más candidatos de manera similar.

Un error común consiste en creer que las poblaciones socialmente diversas siempre serán poblaciones de electores heterogéneas. Sin embargo, simplemente porque una población sea socialmente heterogénea no se deduce que a la hora de votar también lo sea. Por ejemplo, India cuenta con múltiples lenguajes y religiones, pero es relativamente homogénea cuando se trata de construir una muestra de la población votante.

Cuanto mayor sea la heterogeneidad de la población votante, tanto más grande tendrá que ser la muestra a fin de arrojar un estimado preciso del comportamiento electoral. La comparación del tamaño de la muestra requerida para tres países con poblaciones de tamaño sumamente distinto — Canadá, Estados Unidos y Suiza — ilustra este punto.

FIGURA 5-1:
TAMAÑO DESEADO DE LA MUESTRA PARA ESTADOS UNIDOS, CANADÁ Y SUIZA

	ESTADOS UNIDOS	CANADÁ	SUIZA
Población:	263,814,032	28,434,545	7,084,984
Margen de error :	+/-2%	+/-2%	+/-2%
Tamaño deseado de la muestra:	1200	2400	4300

La estrategia más segura es asumir el supuesto conservador de que la población votante es heterogénea.

Como muestra la Figura 5-1, la heterogeneidad no está determinada por las características étnicas de estas poblaciones. Ella está determinada por la probabilidad de que un candidato obtenga la mayoría del respaldo electoral. En un sistema bipartidista como el de los Estados Unidos, la carrera electoral es a menudo más fácil de seguir y mucho más fácil de predecir: los votantes generalmente no tienen más de dos opciones. Pero en Suiza, el mayor número de partidos hace que la competencia electoral sea más complicada. Los partidos políticos suizos están claramente respaldados por grupos de distinto lenguaje y religión. Incluso un país como Canadá, con cinco partidos oficiales, es menos heterogéneo que Suiza.

La Figura 5-1 también ilustra un principio afín. El tamaño de la muestra requerida está determinado por el nivel esperado de homogeneidad en los resultados de la votación, no por la población total de un país. Estos tres países, con poblaciones totales sumamente distintas, requieren muestras de distinto tamaño para conservar un margen de error de más o menos 2% (+/-2%). De hecho, resulta que el país con la población más grande requiere la muestra más pequeña. En realidad, las variaciones en el tamaño de muestra requerido son atribuibles a las variaciones en la homogeneidad de las tres poblaciones distintas.

En la práctica es difícil encontrar información confiable sobre la heterogeneidad u homogeneidad de la población votante de muchos países. En estas circunstancias, la estrategia más segura y que no requiere de adivinanzas, es asumir el supuesto conservador de que la población votante es heterogénea. Como quedará claro luego, este supuesto tiene un profundo impacto sobre cómo se calcula el tamaño de una muestra de conteo rápido.

Cuanto mayor sea la confianza necesaria para que la distribución de la muestra refleje la distribución de la población, tanto más grande deberá ser la muestra.

Niveles de Confianza: Especificando la Relación entre Muestra y Población

Otro tipo adicional de información tiene un importante impacto sobre cómo los estadísticos estiman la población en base a una muestra: el nivel de confianza. Los niveles de confianza se refieren a cómo los datos de la muestra pueden compararse con la población. Cuanto más sea necesario confiar en que la distribución de la muestra refleje la distribución de la población, tanto más grande deberá ser la muestra. Esto se debe a que, en las muestras más grandes, los resultados excepcionales tendrán un efecto menor sobre la distribución.

Para los estadísticos, la práctica convencional es fiarse de un nivel de confianza de 95%. Técnicamente, el nivel de confianza expresa — como porcentaje — la probabilidad con la cual uno está seguro de que la media de la muestra proporcionará un estimado preciso de la media de la población. De este modo, un nivel de confianza del 95% indica que el 95% de todas las medias de la muestra corresponderá realmente a la media de la población. Dado que las consecuencias de un resultado inexacto del conteo rápido pueden ser muy serias, la práctica acostumbrada en la observación de las elecciones es diseñar la muestra con parámetros más conservadores, un nivel de confianza del 99%.

Dado que las consecuencias de un resultado inexacto del conteo rápido pueden ser muy serias, la práctica acostumbrada en la observación de las elecciones es diseñar la muestra con un nivel de confianza más conservador, del 99%.

CÓMO CONSTRUIR LA MUESTRA

La tarea práctica de construir una muestra para el conteo rápido involucra realizar una combinación de selecciones. Entre ellas tenemos:

- identificar la unidad de análisis;
- determinar el margen de error y los niveles de confianza;
- determinar el tipo más apropiado de muestra aleatoria; y
- estimar los factores de corrección para las tasas de recuperación de las muestras y la abstención.

La Unidad de Análisis

La unidad de análisis se refiere al objeto preciso que está siendo examinado. Si el objetivo es generalizar sobre toda una población, entonces la unidad de análisis a menudo es la persona individual. Sin embargo, en algunos casos es posible generalizar una población a partir de una muestra adoptando un agregado más grande como unidad de análisis, tal como una unidad doméstica o una manzana de una ciudad.

En el conteo rápido, el objetivo es estimar la distribución del voto de los ciudadanos entre los partidos políticos. En una elección democrática, el voto

El conteo rápido utiliza por lo general el resultado oficial en una mesa de votación individual como unidad de análisis.

individual es secreto y por ello no puede ser la unidad de análisis. En vez de ello, el conteo rápido utiliza por lo general el resultado oficial en una mesa de votación individual como unidad de análisis.

El Margen de Error: ¿Cuán Exactos Necesitamos Ser?

El margen de error es una de las partes más importantes de la información considerada al construir una muestra. Expresado como un porcentaje, el margen de error se refiere al rango probable de valores para cualquier observación. El siguiente ejemplo ilustra el concepto:

> Los resultados de una mesa de votación indican que el 48% de los votos apoyan al Candidato A. Si el margen de error diseñado es de 5%, hay buenas razones para confiar en que el resultado real del Candidato A caerá en algún lugar entre 43 y 53 por ciento, al considerar a todos los votantes de la población.

Las organizaciones cívicas que efectúan un conteo rápido generalmente diseñan las muestras para que tengan un margen de error de más o menos 0.5% (+/-0.5%).

Las organizaciones cívicas que efectúan un conteo rápido generalmente diseñan las muestras para que tengan un margen de error de más o menos 0.5% (+/- 0.5%). Ocasionalmente hay razones (por ejemplo, la expectativa de que una votación será muy disputada) para seleccionar un margen de error aún más riguroso. El margen deseado depende del nivel de exactitud que se requiera del estimado.

El margen de error se calcula empleando la siguiente fórmula:

$$ME = \frac{s}{\sqrt{n}} * z$$

Donde

ME = margen de error

s = desviación estándar (asúmase 0.5%)

n = tamaño de la muestra

z = valor z para el nivel de confianza escogido (para 95% es 1.96, para 99 es 2.58)

Como todo conjunto de datos, un conjunto de observaciones de puntos de muestra tiene por lo menos dos propiedades. Los datos tendrán una tendencia central, alrededor de la cual se concentrarán la mayoría de los resultados. También tendrán una varianza o dispersión. La varianza se refiere a qué tan amplia o estrechamente están dispersas las observaciones. Hay distintas formas de medir la tendencia central y la dispersión, y estas son relevantes para el cálculo del margen de error.

Medición de la Tendencia Central

La medida más conocida de la tendencia central es la media. La media aritmética es simplemente el valor promedio de todas las observaciones registradas. Se le deriva sumando los valores de cada observación en un conjunto de datos y luego dividiendo entre el número de observaciones. El siguiente ejemplo ilustra este proceso:

> El siguiente conjunto de números: 1, 3, 4, 6, 7 y 9 tiene un promedio de 5. Esto es porque 1 + 3 + 4 + 6 +7 +9 = 30, el número de observaciones es 6, y 30 ÷ 6 = 5.

Hay otras formas de medir la tendencia central de cualquier dato. Por ejemplo, la moda se refiere al número que aparece con mayor frecuencia en cualquier conjunto de datos. En el siguiente conjunto de números: 1, 3, 3, 3, 5, 6 y 7, la observación que se registra con mayor frecuencia es 3. Nótese, sin embargo, que la media aritmética de este mismo conjunto de números es 4 [(1 + 3+ 3+ 3+ 5 +6 +7) ÷ 7 = 4].

Una tercera medida de la tendencia central es la mediana. Este número aparece en el medio de un conjunto dado de observaciones. Para el siguiente conjunto de datos: 1, 3, 6, 7, 8, 8 y 10, el número al centro de las observaciones (mediana) es 7; hay tres observaciones menores que 7 y tres con valores más grandes. Sin embargo, la moda de este conjunto de datos es 8 porque éste aparece con mayor frecuencia. La media aritmética es 6.14. Los estadísticos por lo general reportan la media antes que la mediana o la moda como la medida más útil de la tendencia central.

Medición de la Dispersión

Una segunda característica de los datos se refiere a la medición de la dispersión, que indica qué tan amplia- o estrechamente están distribuidos los datos observados. A partir del ejemplo anterior es claro que todo conjunto de datos dados tendrá una media aritmética. Sin embargo, dicha media no da información alguna sobre qué tan amplia o estrechamente están dispersos los valores observados. Los siguientes conjuntos de datos tienen la misma media aritmética de 3:

> 2, 2, 3, 4, 4
> -99, -99, 3, 99, 99

Estos dos conjuntos de datos tienen distribuciones sumamente distintas. Una forma de expresar la diferencia en los dos conjuntos de datos consiste en considerar el rango de los números. En el primer conjunto, el más pequeño es 2 y el más grande es 4. Entonces, el rango resultante es 4 menos 2, o 2. En el segundo conjunto, el número más pequeño es 99 negativo y el más grande es 99 positivo. El rango resultante es 99 positivo menos 99 negativo, o 198.

Obviamente, los distintos rangos de estos dos conjuntos de datos captan un aspecto de las diferencias fundamentales entre estos dos conjuntos de números. Aun así, al rango solamente le interesan dos números: el más grande y el más pequeño; ignora todos los puntos de datos restantes. Se puede expresar mucha más información sobre la dispersión de las observaciones dentro del conjunto de datos con una medida distinta, la varianza .

En términos laicos, la varianza expresa el promedio de todas las distancias entre cada valor de observación y la media de todos los valores de observación. La varianza toma en cuenta la media aritmética de un conjunto de datos, además de cada uno de los puntos de datos mismos. En consecuencia, ella incluye toda la información necesaria para explicar la dispersión de un conjunto de datos. La varianza de todo conjunto de observaciones puede determinarse con cuatro pasos:

1. Calcúlese la media aritmética del conjunto de datos.
2. Calcúlese la distancia entre cada punto de dato y la media, y elévese la distancia al cuadrado.
3. Súmese todas las distancias elevadas al cuadrado.
4. Divídase esto por el número de observaciones.

Entonces, esta fórmula es como sigue:

For a dataset containing observations x_1, x_2, x_3 … x_n

$$s^2 = \frac{(x_1\text{-}x)^2 + (x_2\text{-}x)^2 + (x_3\text{-}x)^2 \ldots (x_n\text{-}x)^2}{n\text{-}1}$$

Donde
s2 = varianza
x1, x2, x3 … xn son las observaciones
x es la media
n es el número de observaciones

En forma abreviada aparece como sigue:

$$s^2 = \frac{\sum (x\text{-}x)^2}{n\text{-}1}$$

La desviación estándar es la raíz cuadrada de la varianza. Los estadísticos usualmente confían en la desviación estándar porque ella expresa la varianza en unidades estandarizadas que pueden compararse significativamente. Cuanto más grande sea la desviación estándar de cualquier conjunto de datos, tanto más estos estarán dispersos de la media. Cuanto más pequeña sea la desviación estándar, tanto más concentrados estarán los puntos de datos alrededor de la media.

Hay otro concepto de medición adicional que debe considerarse: la distribución normal. La discusión anterior muestra que, en cada conjunto de datos, los puntos de datos individuales se concentrarán en torno a un punto promedio, o media. Otra forma de expresar la misma idea es considerar qué proporción de todas las observaciones cae dentro de una desviación estándar de la media. Si los conjuntos de datos son lo suficientemente grandes y cumplen los principios de aleatoriedad, la dispersión del valor de los datos seguirá lo que se llama una distribución normal. La distribución normal tiene propiedades bien conocidas: la curva normal, como se ve en la Figura 5-2, tiene forma acampanada y simétrica, y la media, el modo y la mediana coinciden.

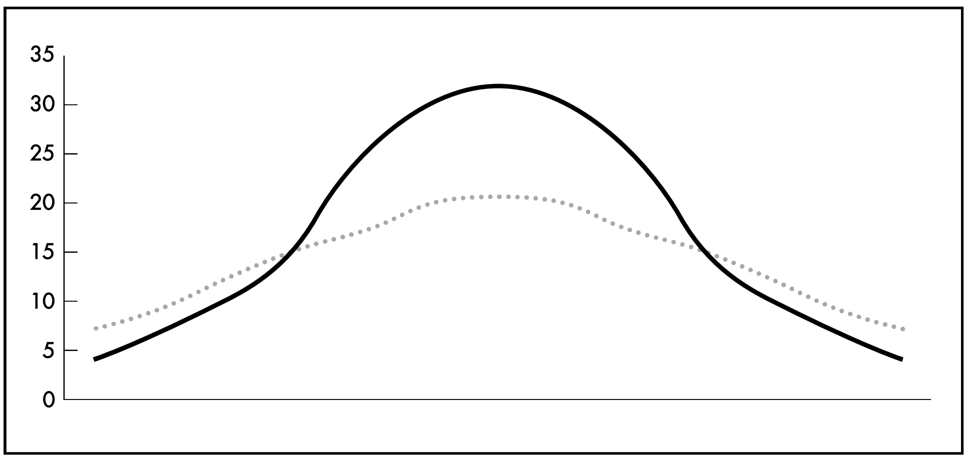


FIGURA 5-2: CURVAS DE DISTRIBUCIÓN NORMALES

El tamaño de la varianza determina la forma precisa de la distribución misma. El punto clave para los fines del conteo rápido es que todo conjunto de datos que sigue la curva de distribución normal tiene las mismas propiedades estándar. Estas son: el 68.3% de todos los valores observados caerán dentro de una desviación estándar de la media, 95.4% de todos los resultados caerán dentro de dos desviaciones estándar de la media, y 99.7% de todos los resultados caerán dentro de tres desviaciones estándar de la media. No todo conjunto de datos coincidirá con este patrón exacto; si hay bastante varianza dentro de los datos la curva será relativamente plana: Si hay poca, la curva tendrá un pico mayor.

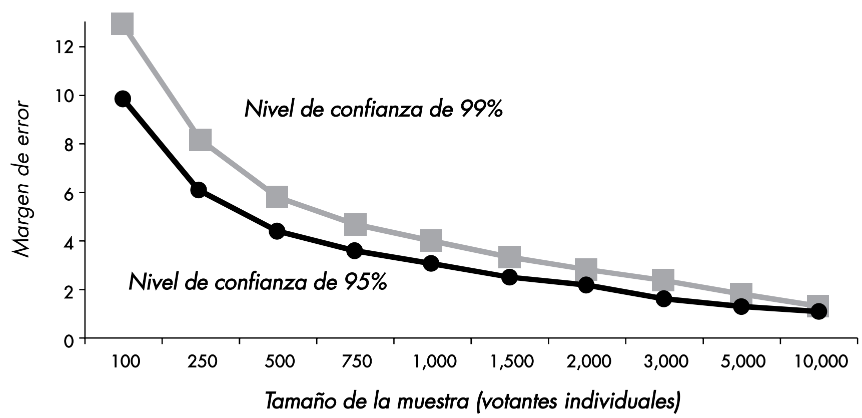
La distancia de la media, expresada como una desviación estándar, puede también ser denominada como los resultados Z o valores críticos. La mayoría de los libros de texto de estadística incluyen una tabla de valores de Z para la distribución normal, y los analistas no tienen que calcularlos cada vez que tienen que vérselas con un conjunto de datos. Significativamente, si los datos

Las decisiones tomadas sobre qué margen de error puede tolerarse en un conteo rápido tendrán un impacto directo sobre los cálculos para determinar el tamaño mínimo requerido de la muestra.

tienen un intervalo de confianza de 95% (el 95% de todas las medias de la muestra incluirá la media poblacional), entonces es claro que los resultados caerán dentro de una desviación estándar de la media de 1.96. De igual modo, un nivel de confianza del 99% indica que el 99% de todos los resultados (para los cuales la media de la muestra incluye la media poblacional) caerá dentro de una desviación estándar de la media de 2.58. En estos casos, los valores 1.96 y 2.58 representan los valores críticos, o valores Z, para los niveles de confianza de 95% y 99%, respectivamente.

Calcular el margen de error requiere depender de la desviación estándar y los valores Z. Estos dos a su vez comprenden la medición de la tendencia central, la medición de la dispersión y los niveles de confianza. Como lo muestra la Figura 5-3, el margen de error varía con el nivel de confianza y el tamaño de la muestra. En general, cuanto más alto sea el nivel de confianza, tanto más alto será el margen de error, y cuanto más grande sea el tamaño de la muestra, tanto más bajo será el margen de error. Las decisiones tomadas sobre qué margen de error puede tolerarse en un conteo rápido tendrán un impacto directo sobre los cálculos para determinar el tamaño mínimo requerido de la muestra.

FIGURA 5-3: EL MARGEN DE ERROR Y EL TAMAÑO DE LAS MUESTRAS



Tipos de Muestras

Hay dos tipos fundamentales de muestras: las muestras probabilísticas y las no probabilísticas. Las primeras cumplen con el principio de la aleatoriedad y son, por lo tanto, representativas de la población total. El conteo rápido siempre usa muestras probabilísticas.

Las muestras no probabilísticas no seleccionan los puntos de muestreo al azar y se ignora el grado en que son representativas de la población más amplia. Estas muestras son útiles bajo ciertas circunstancias. Es más barato y más fácil construirlas y llevarlas a cabo que una muestra probabilística. Los casos de la muestra son elegidos según lo fácil o cómodo que resulta estudiarlos. Por ejemplo, un reportero de televisión se para afuera de un parque de baseball y pregunta a los aficionados si gozaron del juego. La estrategia provee secuencias rápidas e interesantes para la transmisión, pero no brinda información confiable sobre la población total dentro del parque.

Para los fines del conteo rápido, la limitación fatal de las muestras no probabilísticas es que no son confiables para generalizarlas a toda la población. Por lo tanto, la información que producen no es un estimado confiable de las características de la población. Si, por ejemplo, una muestra del conteo rápido se construye íntegramente a partir de mesas de votación en la ciudad capital, los resultados serán casi con toda seguridad distintos que aquellos que provienen de una muestra de mesas de votación en las áreas rurales. Las personas que extraen datos en bruto en lugares convenientes y fácilmente accesibles no están empleando datos representativos de la población como un todo.

Un conteo rápido siempre debe usar muestras probabilísticas para producir así resultados que son representativos de una población definida.

PREGUNTAS FRECUENTES

¿Pueden los estadísticos diseñar muestras para el conteo rápido que combinen las ventajas de la muestra probabilística con los de una muestra no probabilística?

Rotundamente no. Una muestra es o bien una muestra probabilística, en cuyo caso se ciñe íntegramente a los principios de aleatoriedad, o es una muestra no probabilística. Combinar elementos de ambas técnicas de muestreo produciría una muestra no probabilística. En consecuencia, debe evitarse la tentación de sustituir mesas de votación inaccesibles por otras "convenientes". Los puntos de muestreo no pueden sustituirse entre sí porque esa estrategia viola las premisas de la aleatoriedad: que cada punto de la muestra tiene exactamente las mismas probabilidades de ser seleccionado. En vez de ello, los organizadores del conteo rápido deben encontrar una forma de enviar observadores preparados a las mesas de votación contenidas en la muestra, y recibir la información procedente de éstas, incluso si se encuentran en zonas remotas. Reunir y reportar los datos de una mesa de votación situada más convenientemente puede comprometer la confiabilidad y validez de todo el conteo rápido.

Un conteo rápido siempre debe usar muestras probabilísticas para producir así resultados que son representativos de una población definida. Hay varios tipos de muestras probabilísticas y cada uno de ellos puede ofrecer representaciones precisas de la población dependiendo de distintos métodos. Los dos tipos más comunes de muestras probabilísticas son la muestra general aleatoria y la muestra estratificada aleatoria.

El muestreo para un conteo rápido puede comenzar solamente cuando se cuenta con una lista exacta y exhaustiva de todas las mesas de votación.

Muestras Generales Aleatorias

En la muestra general aleatoria, las unidades de análisis se eligen al azar, una por una, de entre toda la población. Esto da a cada unidad de la población una posibilidad igual de ser incluida en la muestra. Sin embargo para que cada unidad de análisis tenga igual posibilidad de ser incluida en la muestra, debe contarse con una lista exacta de todas las unidades posibles de análisis.

Los estadísticos llaman marco del muestreo a la lista de todos los miembros de la población. En el caso de un conteo rápido, la unidad de análisis es la mesa de votación; por lo tanto, el muestreo para un conteo rápido solamente puede comenzar cuando se cuenta con una lista exacta y comprehensiva de todas ellas.

Muestras Estratificadas Aleatorias

La muestra estratificada aleatoria aplica el mismo principio de aleatoriedad que la muestra general aleatoria. Sin embargo, los marcos del muestreo a partir de los cuales se eligen los puntos de la muestra constan de estratos predeterminados y mutuamente excluyentes de la población total. Por ejemplo:

> El objetivo de un proyecto es utilizar una muestra de 1,000 estudiantes para generalizar sobre una población universitaria de 20,000 alumnos, la mitad de los cuales son alumnos de postgrado. Mientras que el método general de la muestra aleatoria simplemente elige al azar 1,000 puntos de muestra de la lista total de 20,000 estudiantes, el método de la muestra estratificada sigue dos pasos. Primero, se divide la lista de todos los estudiantes en dos grupos (estratos), uno que incluye a todos los alumnos de pre-grado y el otro a todos los de postgrado. Luego se seleccionan 500 casos del estrato 1 (pre-grado) y otros 500 del estrato 2 (postgrado).

La práctica de la estratificación significa que el resultado final arrojará una muestra total que refleja ***exactamente*** la distribución de casos en la población como un todo.

En el método estratificado, la selección de cada caso continúa satisfaciendo el criterio de aleatoriedad: la probabilidad de la selección de cada caso dentro de cada estrato es exactamente la misma (en el ejemplo anterior, 1 de 20). Sin embargo, la práctica de la estratificación significa que el resultado final arrojará una muestra total que refleja *exactamente* la distribución de casos en la población como un todo. En efecto, el procedimiento de estratificación predetermina la distribución de casos en todos los estratos.

La estratificación podría ser útil de otro modo. Algunos grupos de observadores no cuentan con los recursos para llevar a cabo una observación a nivel nacional. En ese caso, el grupo podría limitar su observación a un estrato particular del país, tal vez la ciudad capital, o una región costera. En estos casos, con un conjunto seleccionado al azar de puntos de muestreo dentro de un estrato, el grupo de observadores puede generalizar los resultados de la observación a todo el estrato que el grupo de observación cubre.

Cómo Determinar el Tamaño de la Muestra

Para determinar el tamaño de la muestra para un conteo rápido (esto es, cuántas mesas de votación deben incluirse en ella), los analistas efectúan una serie de procedimientos. Identifican el tamaño de la población relevante (el número de votantes elegibles); determinan el nivel de homogeneidad dentro de esa población y seleccionan el nivel de confianza y el margen de error deseados.

PREGUNTAS FRECUENTES

¿Dado que la mayoría de las poblaciones parecen estar estratificadas cuando se trata del voto, ¿por qué no usar muestras estratificadas como una cuestión de rutina?

El método de estratificación parece ser ideal, pero hay dos razones por las cuales podría no ser apropiado. En primer lugar, si el conteo rápido está utilizando supuestos conservadores sobre los márgenes de error y los intervalos de confianza, entonces es probable que la muestra sea demasiado grande. Y debido a la teoría de la probabilidad, es claro que las muestras grandes terminan produciendo réplicas precisas de la población total incluso sin la estratificación. Segundo, la estratificación asume información confiable sobre cómo los ciudadanos tienden a votar dentro de los estratos. ¿Exactamente alrededor de qué lineamientos deben trazarse los criterios de estratificación? En muchos países, la información necesaria para efectuar ese juicio podría no ser muy confiable. Las elecciones anteriores no pueden ser una guía confiable, en particular si sus resultados fueron cuestionados. La estrategia más sólida es referirse a estratos *post facto*, o sea, contrastar la distribución de los casos extraídos de una muestra general aleatoria con estratos de la población después de prepararse la muestra. De modo tal que si el 40% de la población votante vive en la ciudad capital de un país, entonces el 40% de los puntos de muestra extraídos al azar deberían corresponder a la ciudad capital.

Luego los analistas calculan el tamaño de la muestra como sigue:

$$n = \frac{P\,(1-P)}{\frac{\Sigma^2}{z^2_{99\%}} + \frac{P\,(1-P)}{N}}$$

Donde

n = tamaño de la muestra (el número de votantes en regla)

P = el nivel sospechado de homogeneidad de la población (entre 0 y 1, de manera que 50% = 0.5)

Σ = margen de error (entre 0 y 1, de modo que 0.32% = 0.0032)

$z_{99\%}$ = nivel de confianza en el caso de una distribución normal (99% en este caso)

N = tamaño de la población total

El caso de un conteo rápido efectuado en las elecciones presidenciales del 2001 en Perú puede ilustrar los pasos anteriores:

> El tamaño de la población total relevante (el número de votantes en regla) en Perú era de 14,570,774. Se asumió que la población era heterogénea: se esperaba que la contienda entre ambos candidatos fuese muy reñida, de modo que el nivel de homogeneidad de la población se fijó en 50% (0.5). Se eligieron un margen de error de 0.32% y un nivel de confianza de 99%. Para los fines de hacer el cálculo se expresó la proporción de homogeneidad como un valor con un rango entre 0 y 1, al igual que el margen de error. El nivel esperado de homogeneidad se fijó en 50%, el supuesto más conservador; en la fórmula aparece como 0.5%, y el margen de error de 0.32% (de un 100% posible) se expresa como .0032. Estos valores se colocaron en la fórmula como sigue:

$$N = \frac{.50\,(1-.50)}{\dfrac{(.0032)^2}{(2.58)^2} + \dfrac{.50\,(1-.50)}{14{,}570{,}774}}$$

$$= \frac{.25}{\dfrac{.000010}{6.6564} + \dfrac{.25}{14{,}570{,}774}}$$

$$= \frac{.25}{.000001515 + .000000017}$$

$$= \frac{.25}{.000001532}$$

$$= \mathbf{163{,}185}$$

Ya entonces los analistas saben cuántos votantes deben consultarse. Sin embargo, las unidades de análisis no son los votantes individuales, sino mesas de votación. Por lo tanto, el siguiente paso es determinar cuántas de estas mesas deben seleccionarse para que representen el número requerido de votantes. Podemos utilizar el cálculo peruano para ejemplificar este punto:

> En promedio hay aproximadamente 160 votantes por mesa de votación en Perú. Por lo tanto, el tamaño de la muestra de 163,185 (votantes elegibles) fue dividido entre el número de electores por mesa (160) para así determinar el número de mesas de nuestra muestra (1,020). En consecuencia, el tamaño de la muestra para el conteo rápido en Perú el año 2001 era de 1,020 mesas de votación.

Cómo Seleccionar los Puntos de Muestra

Una vez que se conoce el tamaño de la muestra aleatoria, puede seleccionarse la muestra a partir del marco de la misma. Para el conteo rápido se seleccionan las mesas de votación (los puntos de muestra) de la lista completa de mesas (el marco de la muestra). La forma más simple de hacer esto es utilizando un programa de computación aleatorio. Sin embargo, también puede hacerse sin computadora. El primer paso comprende la división del número total de mesas de votación por el número deseado de las mismas, y el segundo paso requiere que se determine un punto de partida aleatorio. Una vez más podemos usar las cifras del conteo rápido peruano de 2001 para mostrar cómo se hace esto:

> El día de las elecciones, el universo peruano constaba de 90,780 mesas de votación. Primero se dividió el número total de las mismas por el número deseado de mesas en la muestra (90,780 ÷ 1,020 = 89). Esto indica que debía seleccionarse una de cada 89 mesas. En segundo lugar se eligió un punto de partida aleatorio colocando 89 tiras de papel en un sombrero, numeradas del 1 al 89, y eligiendo una de ellas al azar. La tira seleccionada tenía el número 54. La 54ª mesa de votación de la lista ordenada al azar era el primer punto de muestra, eligiéndose luego cada 89ª mesa después del primer punto. De este modo, la segunda mesa de votación de la muestra fue la 143ª de la lista (54 más 89). Se repitió el procedimiento hasta alcanzar el tamaño total de la muestra de 1,020.

El ordenamiento al azar es una técnica que proporciona seguridad adicional de que la probabilidad de selección de cada punto de la muestra es igual a la que tiene cualquier otro punto de ser elegido.

¿Por qué la lista de mesas de votación debe ordenarse al azar? Esta estrategia protege aún más la validez y confiabilidad del conteo rápido. Si la lista original está organizada por tamaño, región u otros criterios, el resultado de un sorteo simple podría estar sesgado. Por lo general este no es un problema serio, pero el ordenamiento al azar es una técnica que proporciona seguridad adicional de que la probabilidad de la selección de cada punto de la muestra es igual a la de que se elija cualquier otro punto.

Factores de Corrección

A veces es necesario efectuar ajustes a diversos elementos de la metodología del conteo rápido. Estos ajustes se aplican al reclutamiento y la capacitación de los voluntarios, y a elementos más técnicos del conteo rápido, incluyendo el muestreo. El cálculo de la muestra arriba esbozado por lo general requiere de ciertos ajustes adicionales. Esto se debe a que inicialmente se asume que todos los puntos de muestra serán identificados y que los datos se entregarán desde todos y cada uno de ellos. Sin embargo, en la práctica, ningún conteo rápido a gran escala, emprendido por cualquier grupo de observación, ha logrado entregar la información de todos los puntos de muestra en la muestra original.

Ningún conteo rápido a gran escala emprendido por cualquier grupo de observación ha logrado entregar la información de todos los puntos de muestra en la muestra original.

En la situación de un conteo rápido resulta importante trazar una distinción entre una muestra teórica y una muestra práctica. La mayoría de las discusiones teóricas del muestreo asumen que, una vez elegido un punto de muestra, la

En una elección sumamente disputada, los datos faltantes pueden constituir un serio problema.

información del mismo se generará con una eficiencia del 100%. Este supuesto jamás se ha cumplido en ningún conteo rápido nacional a gran escala. Ello se debe a diversas combinaciones de factores, entre ellos los errores cometidos por observadores capacitados en forma inadecuada, las rupturas en el sistema de comunicaciones, o eventos inesperados el día de la votación. (Por ejemplo, a veces se prohíbe a los observadores ingresar a los centros de votación; un clima inclemente puede impedir que los observadores lleguen a un teléfono, o que los datos sean reportados.)

En promedio, las organizaciones cívicas que efectúan un conteo rápido por primera vez logran entregar alrededor del 75% de los datos de los puntos de muestra dentro de un marco temporal razonable, unas 3 horas. El 25% de la muestra que no se reporta (estos son datos faltantes) puede causar problemas para la interpretación de los demás datos. Por lo tanto, la muestra práctica utilizable es siempre más pequeña que la muestra diseñada teóricamente. El margen de error que se aplica a la muestra práctica debe asimismo ser mayor de lo originalmente planeado.

En una elección sumamente disputada, los datos faltantes pueden constituir un serio problema. Es más, estos datos rara vez constituyen un corte trasversal aleatorio de la muestra total. En la práctica, la proporción de datos faltantes casi siempre es mayor en las zonas remotas donde son más difíciles de recuperar. Los datos faltantes están sesgados si no son aleatorios o representativos. Y si los datos faltantes están sesgados, entonces el resto de la muestra también lo estará.

¿Cuál es la mejor forma de prepararse para el hecho de que no toda la muestra se recuperará el día de las elecciones? La respuesta debe incluirse en el diseño original de la muestra; se trata de sobre-muestrear según el margen de la tasa de recuperación esperada.

El estadístico debe sobre-muestrear según el margen de la tasa esperada de recuperación.

Un grupo de observación experimentado podría contar con una tasa de recuperación esperada del 80% de los puntos de muestra de la muestra teórica. En este caso, la muestra práctica sería 20% más pequeña que la teórica. La forma más directa de hacer frente a este problema potencial es simplemente incrementando el tamaño de la muestra en 20%, agregando aleatoriamente 20% más puntos de muestra que lo calculado inicialmente. Una estrategia directa como esta funciona si el déficit en la recuperación de la muestra está distribuido aleatoriamente por toda la población. Sin embargo, la experiencia indica que el déficit generalmente está repartido de manera desigual entre la ciudad capital, otras áreas urbanas y las áreas rurales. La dificultad mayor yace en las zonas remotas, y el diseño de un componente corregido de sobre-muestreo debe tener esto en cuenta. La Figura 4-5 muestra la distribución de un típico patrón de recuperación de la muestra y el componente corregido de sobre-muestreo. Como indica la Figura 5-4, la corrección adicional para la recuperación desigual de la muestra coloca por lo menos la mitad del sobre-muestreo en las zonas rurales.

PREGUNTAS FRECUENTES

¿Acaso la muestra no es ya lo suficientemente grande, puesto que son las mesas de votación las que están siendo muestreadas, y no los votantes individuales? En ese caso, ¿no pueden los analistas simplemente ponderar los datos después de recogida la muestra?

Si, la muestra es grande tal como se la diseña, pero ponderar los datos no reemplaza los datos reales. La ponderación simplemente da "más peso" a los datos existentes de la muestra global. No hay forma de saber si los datos faltantes de las áreas remotas de la muestra son típicos de los datos recuperados en esa sub-muestra particular. La ponderación es una estrategia estadística usada como último recurso luego de agotarse todas las demás opciones.

	RECUPERACIÓN DE LA MUESTRA	DISTRIBUCIÓN DE LA SOBREMUESTRA
Ciudad capital:	85%	15%
Áreas urbanas afuera de la ciudad capital	75%	25%
Áreas remotas	65%	35%

FIGURA 5-4: UN TÍPICO PATRÓN DE RECUPERACIÓN DE LA MUESTRA Y LA DISTRIBUCIÓN RECOMENDADA DEL SOBRE-MUESTREO

Cómo Hacer Correcciones Según el Tamaño de las Mesas de Votación

Por razones prácticas, a veces es necesario ajustar el margen de error para los resultados del conteo rápido. Por ejemplo, el tamaño de la mesa de votación —el número de votantes esperado en estas mesas— afectará el margen de error. Esto se debe a la diferencia entre la población definida y la unidad de análisis. Recordemos que el cálculo original del margen de error dependía del número total de votantes en regla. Esto se hizo para que el diseño de la muestra cumpliera con ciertos principios de la estadística. Sin embargo, como las mesas de votación son las unidades de análisis, es útil revisar el margen de error en base al número de votantes en ellas. En el ejemplo anterior, un promedio de 160 votantes fue asignado a cada mesa de votación. Habría sido importante considerar el hecho de que estas mesas pueden tener distinto tamaño. Si comprenden 200 votantes, ello habría tenido el efecto de reducir el número de mesas necesarias para la muestra. Si son aún más grandes, con 500 votantes, entonces aún menos serían necesarias para constituir la muestra.

El número de mesas de votación y el de votantes en una mesa tendrá un efecto sobre el margen de error.

Como lo muestra la Figura 5-5, el número de mesas de votación y el de votantes en una mesa tendrá un efecto sobre el margen de error. Esto se puede atribuir al papel que tiene el tamaño de la muestra en la construcción del margen de error. Recuérdese que la fórmula de este último es:

$$\frac{\text{(Heterogeneidad asumida) * (valor de z en el nivel de confianza elegido)}}{\sqrt{n}}$$

El hecho es que la variación en el tamaño de las mesas de votación también afectará a 'n'.

FIGURA 5-5:
TAMAÑO DE LA MUESTRA Y MÁRGENES DE ERROR

	VOTANTES	MESAS DE VOTACIÓN		
		Si el tamaño de la mesa es 160	Si el tamaño de la mesa es 200	Si el tamaño de la mesa es 500
Muestra	163,185	1,020	816	324
Margen de error (nivel de confianza 95%)	±0.24	±3.01	±3.43	±5.4
Margen de error (nivel de confianza 99%)	±0.32	±4.03	±4.5	±7.1

Nótese que el margen de error depende de la cantidad de las mesas de votación de la muestra. Si su tamaño es grande, se requerirán menos para generar la muestra deseada de 163,185 votantes. El margen de error calculado para las mesas de votación es menor que aquel calculado para la muestra de votantes. El margen de error resultante para el conteo rápido cae en algún lugar entre los márgenes de error más bajo y más alto.

El margen de error se incrementa a medida que disminuye el número de mesas de votación necesarias para formar una muestra de los votantes.

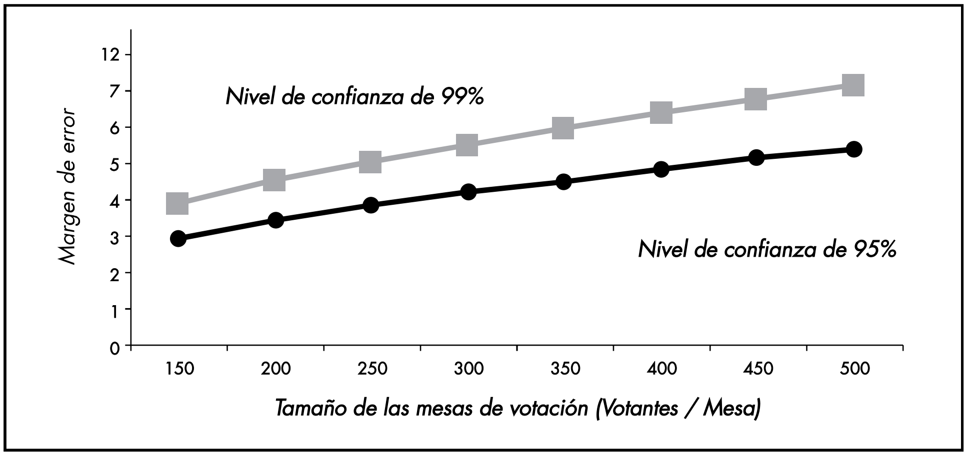
Rastrear los cambios hasta el margen de error para un rango del tamaño de las mesas de votación muestra que este margen se incrementa a medida que disminuye el número de mesas necesarias para formar una muestra de los votantes.

FIGURA 5-6:
TAMAÑO DE LAS MESAS DE VOTACIÓN Y MARGEN DE ERROR

VOTANTES / MESA	# DE MESAS PARA ALCANZAR LA MUESTRA	MARGEN DE ERROR (niveles de confianza) 95%	99%
150	1,088	±2.97	±3.91
200	816	±3.43	±4.52
250	653	±3.84	±5.05
300	544	±4.20	±5.53
350	466	±4.54	±5.97
400	408	±4.85	±6.39
450	363	±5.14	±6.77
500	327	±5.42	±7.13

El margen de error se incrementa a medida que crece el tamaño de las mesas de votación. Sin embargo, el efecto global que su tamaño tiene sobre el margen de error disminuye a medida que ambos se incrementan. La Figura 5-7 ilustra este punto.

FIGURA 5-7:
TAMAÑO DE LAS MESAS DE VOTACIÓN Y MARGEN DE ERROR



Cómo Hacer Correcciones Según la Asistencia

Cuando las elecciones están muy reñidas, los analistas del conteo rápido deben también preocuparse con el nivel de asistencia de los votantes. Aunque los observadores hayan logrado recuperar los datos de cada una de las 1,020 mesas de votación de la muestra teórica, una baja asistencia a las urnas significa que quedan incluidos en la muestra menos votos que si ésta hubiese sido alta. El cálculo original se basaba en la expectativa de unos 160 votos por mesa. Sin embargo, si la asistencia fue de 70%, habrían solamente 112 votos en cada una de ellas. Si este patrón se repite en todas las 1,020 mesas de votación, entonces el conteo sólo incluiría 114,240 votos, unos 50,000 menos de los 163,185 deseados para alcanzar un margen de error de 0.3%, y un nivel de confianza de 99%.

Cuando las elecciones están muy reñidas, los analistas del conteo rápido deben también preocuparse por el nivel de asistencia de los votantes.

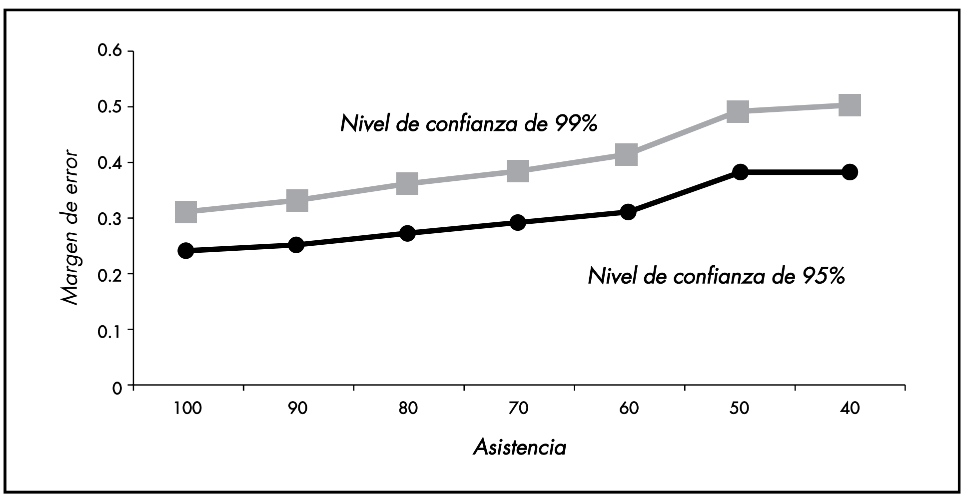
FIGURA 5-8:
ASISTENCIA Y MARGEN DE ERROR

ASISTENCIA	# VOTANTES, VOTOS	MARGEN DE ERROR (niveles de confianza) 95%	99%
Muestra deseada (asistencia=100%)	163,185	±0.24	±0.31
90%	146,867	±0.25	±0.33
80%	130,548	±0.27	±0.36
70%	114,230	±0.29	±0.38
60%	97,911	±0.31	±0.41
50%	81,593	±0.38	±0.49

Entonces, una estrategia cautelosa de interpretación de los datos requiere que se vuelva a calcular el margen de error en base al número real de votantes tabulados. La Figura 5-8 muestra este punto.

Como lo muestra el cuadro, el margen de error se incrementa a medida que la asistencia disminuye. Si esta se encuentra por encima del 60%, el margen de error se incrementará aproximadamente en 0.02% por cada 10 puntos porcentuales menos en la asistencia. Cuando esta se acerca al 50%, el incremento en el margen de error es mucho mayor. La Figura 5-9 muestra un gráfico del incremento en el margen de error correspondiente a la caída en la asistencia.

FIGURA 5-9:
MÁRGENES DE ERROR Y ASISTENCIA



Este capítulo ha presentado a un público general los principios estadísticos generales en que se basa el conteo rápido, y ha esbozado las bases estadísticas de la metodología del mismo. Los organizadores deben entender esta metodología, en especial los conceptos de confiabilidad y validez, así como la razón por la cual una muestra debe cumplir con el criterio de la aleatoriedad. Este conocimiento es vital para el diseño de formularios de observación y programas de capacitación de observadores efectivos y confiables. También hace hincapié en la importancia que tiene preparar la recuperación de los datos de todas las partes del país, incluso las más remotas.

Por último, el capítulo también consideró cuestiones más técnicas de cómo calcular el tamaño de las muestras, y cómo cuestiones tales como los niveles de confianza, los márgenes de error y la heterogeneidad u homogeneidad de la población configuran la muestra. La mayoría de los grupos de observación contratan los servicios de un estadístico calificado para que construya y extraiga una muestra, y analice los datos el día de las elecciones. Los grupos cívicos deben entender que el conteo rápido es una aplicación de principios estadísticos a circunstancias prácticas y singulares en las cuales las suposiciones corrientes de los libros de texto podrían no cumplirse. Por ese motivo, el capítulo esboza cuáles son los factores de corrección más usuales que deben tenerse en cuenta cuando los analistas se enfrentan a la interpretación de los datos recuperados exitosamente el día de las elecciones.

RECUERDE QUE

Las personas que no son estadísticos pueden comprender fácilmente los principios generales en los que se basa el conteo rápido, y hay razones importantes por las cuales el personal clave de los grupos de observación debe familiarizarse con estos principios:

1.Comprender la importancia que tiene asegurar la solidez de los datos del conteo rápido facilitará las decisiones sobre el diseño del mismo y ayudará al personal a preparar formularios de observación y programas de capacitación de observadores eficaces.

2.El personal que aprecie la relación existente entre una muestra y una población, y el lugar central que el requisito de la aleatoriedad tiene para la integridad de dicha relación, estará motivado para construir una robusta red de voluntarios que pueda cubrir hasta las mesas de votación más remotas.

Los grupos deben conseguir el respaldo de un estadístico experimentado en la realización de conteos rápidos para que lleve a cabo las tareas técnicamente complejas de la construcción de una muestra y el análisis de los resultados del conteo rápido. La experiencia con los mismos por todo el mundo subraya varios puntos:

1.La unidad de análisis en un conteo rápido es la mesa de votación. El muestreo no puede comenzar hasta que no se cuente con una lista exacta y detallada de las mesas de votación (el marco del muestreo).

2.El conteo rápido utiliza siempre muestras probabilísticas (por ejemplo, muestras generales aleatorias o muestras estratificadas aleatorias) para producir resultados que sean representativos de toda la población.

3.Los grupos de observación que llevan a cabo el conteo rápido jamás logran recuperar el 100% de los datos de la muestra. Los analistas deben prepararse para lo inevitable. La solución, que puede ser incorporada al diseño original de la muestra, consiste en sobremuestrear según los márgenes de la tasa de recuperación esperada.

4.Los analistas deben asimismo considerar factores de corrección al diseñar una muestra. Los más importantes son aquellos que consideran las variaciones en (a) la asistencia de los votantes y (b) el número de votantes en la unidad básica de análisis: la mesa de votación.