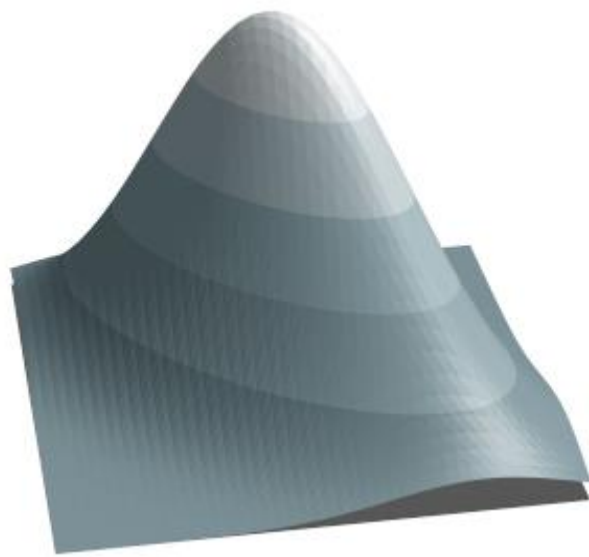


2022

NOTAS DEL CURSO DE MUESTREO PROBABILISTICO PARA ENCUESTAS

Profr. Francisco Sánchez Villarreal



1. GENERALIDADES

1.1 Introducción.

La recolección de información mediante encuestas sobre diversos fenómenos sociales, económicos y políticos se ha popularizado notablemente en todo el mundo. Revistas y periódicos publican diariamente resultados de sondeos de opinión sobre diversos temas de actualidad. Los planificadores y evaluadores de programas educativos, económicos y de salud pública recurren a la encuesta como instrumento fundamental. Los investigadores de mercados y publicistas aplican encuestas para medir la aceptación del público a un nuevo producto o medir el impacto de una campaña publicitaria. En época de elecciones. Instituciones de diversa orientación política realizan encuestas predictivas del voto, encuestas a la salida de las casillas y conteos rápidos sobre resultados electorales. Los organismos responsables de las estadísticas de fenómenos demográficos y macroeconómicos complementan la información recolectada a partir de censos o registros administrativos con datos de encuestas por entrevista directa, teléfono y sitios de Internet.

En las encuestas se utilizan diversas metodologías de muestreos probabilísticos, de cuota o intencionales y ello ha dado origen también a el abuso de prácticas inadecuadas o dolosas que reducen notablemente la confianza del público en sus resultados. La aplicación de encuestas y censos, cuyos orígenes se remontan a los de nuestra cultura es un producto desarrollado fundamentalmente en los últimos 70 años y ha llegado para quedarse como la mejor alternativa para obtener datos específicos e inmediatos de cualquier tema social, demográfico económico. Las innovaciones tecnológicas en metodologías de diseño, instrumentos electrónicos para la recolección y transmisión de datos; paquetería de software especializado para el proceso de encuestas, han facilitado la ampliación de sus aplicaciones.

1.2 Población y Muestra.

El primer paso en la realización de una investigación, consiste en establecer los objetivos que se pretenden alcanzar. Los objetivos se deben plantear tan clara y específicamente como sea posible para establecer los límites del estudio. Ello permitirá hacer una elección adecuada del universo que se someterá a estudio. Desde el punto de vista del muestreo para poblaciones finitas, se entenderá por población o universo cualquier colección de personas u otro tipo de elementos que posean ciertas características en común. Los elementos de la población o universo, son las unidades de las que se busca información y se determinan en función de los objetivos del estudio. Es posible en este sentido hacer referencia a una población de establecimientos comerciales, fábricas de artículos electrónicos, operadoras de vehículos de carga, de predios agrícolas dedicados al cultivos diversos etc.

La definición de la población o universo en estudio se debe precisar en términos de:

El Contenido. Es decir, el tipo de elementos constitutivos, por ejemplo: electores, amas de casa, estudiantes, viviendas, hogares, predios, empresas, etc.

La Extensión. Se refiere a las atribuciones geográficas y/o administrativas asociadas a los elementos, por ejemplo: los estudiantes de la Facultad de Ciencias de la UNAM, las fábricas de prendas de vestir del área metropolitana de la ciudad de México, los predios ejidales del estado de Puebla, los profesores de escuelas de enseñanza media, los productores de aguacate del estado de Michoacán, etc.

El Tiempo. La mayor parte de las poblaciones presentan cambios en función del tiempo. Una encuesta es una medición transversal, vigente durante un período relativamente corto, análoga a una fotografía, por tanto, se debe referir el momento en el que se hace el estudio. Por ejemplo, los estudiantes de la Facultad de Ciencias de la UNAM en el período escolar 2014-II, Las fábricas de prendas de vestir del área metropolitana de la ciudad de México en abril de 2015, etc.

Una muestra es una fracción de la población o universo en estudio, pero para que sea considerada como tal, debe proceder de métodos de selección que incorporen procedimientos de aleatorización, que garanticen su objetividad y representatividad técnica.

1.3 Definiciones Operacionales.

Los aspectos de la población que se pretenden medir conducen a la definición de conceptos y procedimientos de medición. En este punto se suelen presentar frecuentes discrepancias que no son atribuibles a las técnicas de muestreo, sino a las bases conceptuales. Los investigadores suelen definir los conceptos en función de los intereses y alcances de su proyecto, pero es conveniente hacer algunos comentarios para profundizar un poco en este punto.

En 1927 Percy W. Bridgman, físico norteamericano que obtuvo el premio Nobel en 1945, escribió un libro llamado *"Lógica de la Física Moderna"* en el que propuso que los conceptos se definieran por las operaciones con que se acostumbra medirlos. Según él, por ejemplo, *"el concepto de longitud involucra, ni más ni menos, al conjunto de operaciones con que se determina la longitud"*. Significa, por ejemplo, que si medimos algo de dos maneras o instrumentos diferentes, tenemos en realidad dos conceptos diferentes. En los siguientes años surgieron muchos entusiastas y críticos del *"Operacionalismo"*, como se le llamó esta corriente científica. Sus defensores plantean que las definiciones operacionales ayudarían a evitar nociones ambiguas e incluso contradictorias. El campo de aplicación del operacionalismo trascendió a las operaciones físicas y se extendió a operaciones lógicas. El operacionalismo ha tenido sus detractores ante los abusos por el reduccionismo que se puedan hacer en la aplicación de las definiciones, pero sin duda ha permitido que en muchos ámbitos se puedan definir claramente conceptos, no necesariamente con aceptación universal, sino solamente en el consenso de un reducido círculo de interesados en los temas de su especialidad. En la actualidad no se encuentra un protocolo de investigación que no inicie con las definiciones operacionales, materiales y métodos empleados.



Definir en forma clara y precisa es condición previa para posteriormente medir. Por ejemplo, si se desea contar el número de localidades urbanas en una entidad federativa, antes hay que definir el concepto de localidad urbana. La definición operacional de localidad urbana utilizada con mayor frecuencia en los medios oficiales de nuestro país es: "Localidad urbana es aquella que tiene 2500 habitantes o más" ha sido motivo de frecuentes discusiones, pues no refleja los aspectos asociados al urbanismo, como son alumbrado, agua potable, drenaje, pavimento, vigilancia, servicios médicos, servicios educativos, transportes, etc. Así, una localidad rural con 2,499 habitantes puede cambiar a urbana de un día para otro con el solo nacimiento de un bebé.

El lector podrá percibir la dificultad de definir si intenta definiciones operacionales de conceptos tan familiares como: habitación, silla, predio, familia, etc.

Las discrepancias conceptuales perturban frecuentemente los resultados estadísticos, pero desafortunadamente se responsabiliza a la estadística y no a las actividades conceptuales.

1.4 Ventajas y Limitaciones del Muestreo.

Los atributos de una muestra se infieren a la población, ello implica un proceso de generalización que debe apoyarse en técnicas estadísticas para tener una medida del error en que se incurre y una garantía de objetividad.

El estudio de la población completa (censo), en muchas ocasiones resulta impráctico o costoso y ello obliga al análisis mediante muestras. Francisco Azorín Poch, menciona las siguientes **situaciones típicas en el empleo de muestras**:

Cuando la **población** sea tan **grande** que el análisis completo o censo exceda las posibilidades económicas del organismo investigador.

Cuando la población presente **uniformidad en su definición** y una muestra de tamaño adecuado proporcione una buena representación de la misma y por tanto carezca de sentido examinar la población completa.

Cuando el **proceso de medida** o investigación de las características **sea destructivo** o pueda modificar en alguna forma a las unidades observadas.

Por otra parte, se pueden mencionar como las principales ventajas del muestreo, las siguientes:

Bajos costos. Es evidente que la obtención de información de una fracción de la población requiere de menores recursos que los necesarios para un censo, pero no es barata. Es importante destacar que la información por cualquier método que sea recolectada es un producto caro, de utilización limitada y de vigencia muy corta.

Poco tiempo. La recolección y procesamiento de datos de una fracción de la población se realiza en menor tiempo.

Actualidad de datos. Los datos recolectados a partir de una encuesta proceden de un tiempo cercano a su análisis y por tanto las conclusiones de ese análisis tendrán referencias más actuales de la población.

Oportunidad de resultados. Los datos recolectados se utilizan en el momento que son requeridos para su análisis y evitan retrasos en los procesos de producción de resultados.

Desagregación ajustada a necesidades. La desagregación de las estadísticas que se puede lograr a partir de una encuesta puede ser refinada tanto como el investigador lo desee y los recursos disponibles lo permitan.

Mejor capacitación del personal de campo. Las personas que recolectan la información para una encuesta pueden ser capacitadas y supervisadas con más cuidado debido a su reducido número y verificados los datos que recolectan. Todo ello se refleja en notables mejoras en la integridad y precisión de los datos recolectados.

Ante las ventajas señaladas para el muestreo, es conveniente mencionar también sus principales limitaciones:

El riesgo que tienen las muestras de **sufrir serios desvíos** o sesgos motivados por metodologías o prácticas de campo inadecuadas y que desde luego se reflejan en las estimaciones de las características de la población que se desean estimar.

Al trabajar con una fracción de la población **no se puede ser concluyente** en afirmaciones pues siempre existe la probabilidad de errores de muestreo, aunque esa probabilidad de errores importantes sea baja. Adicionalmente hay que considerar los errores no atribuibles al muestreo, cuya medición no suele ser cuantificable fácilmente.

La necesidad de **personal especializado** para diseñar la muestra, efectuar la aplicación de cuestionarios y el proceso estadístico de datos.

Elevadas tasas de rechazo de los probables entrevistados pueden inducir sesgos y en consecuencia restar representatividad a la muestra debido a que los sujetos que aceptan ser entrevistados, no necesariamente tienen las mismas características que los sujetos que rechazan las entrevistas.

1.5 El Muestreo Probabilístico.

El procedimiento por el cual se seleccionan elementos de la población, para su incorporación a la muestra, define el que se trate o no, de un muestreo probabilístico. En un muestreo probabilístico, es posible definir el espacio muestral o conjunto de todas las muestras posibles. Además de que, es posible asociar a cada muestra una probabilidad conocida y distinta de cero de ser seleccionada.

Con el fin de evitar posibles sesgos en el proceso de selección, se incorporan a éste, prácticas de aleatorización que lo hacen independiente al diseñador de la muestra seleccionada.

Finalmente, en el muestreo probabilístico, se define un método de estimación que proporciona un resultado único para cada muestra.

El desarrollo de la teoría de muestreo se refiere al muestreo probabilístico, aunque en la práctica no es la única alternativa pues también se recurre frecuentemente a muestreos intencionales y de cuota, cuya limitante es la frecuente falta de objetividad y el no cumplimiento de los supuestos de las técnicas de inferencia. También hay que considerar situaciones de riesgo en muestras para ensayos clínicos y entonces la muestra es autoselectiva pues se recurre a voluntarios.

1.6 El Marco de Muestreo.

A la definición conceptual de la población objetivo debe corresponder una forma práctica de tener acceso a todos los elementos que la constituyen. En forma simple, la relación de todos los elementos que integran una población o universo de estudio se le conoce como marco de muestreo. Es frecuente el empleo de marcos que hacen referencia a superficies rurales o urbanas, tales como mapas de terrenos de cultivos o planos de traza urbana de localidades. Estas últimas formas de marcos se identifican genéricamente como marcos área.

La disposición de marcos adecuados suele ser una problemática frecuente para el investigador. Los archivos administrativos que usualmente se utilizan como marco de muestreo frecuentemente presentan serias omisiones de elementos de la población objetivo, repeticiones o inclusiones de elementos extraños. Una conducta deseable para el investigador es la de evaluar los problemas que pueda tener un marco mediante la verificación en campo de los datos que aparecen en los registros administrativos. Idealmente construir específicamente un marco sería la mejor opción, pero su costo suele estar fuera de los presupuestos.

1.7 Etapas en el Desarrollo de una Encuesta.

La realización de una encuesta involucra una serie de actividades de planeación y organización cuya complejidad puede variar notablemente. Por ejemplo, una encuesta de opinión entre los estudiantes de una preparatoria de 1200 alumnos, supone una situación muy diferente a una encuesta de producción agropecuaria en todo el país. Sin embargo, existen una serie de actividades que podríamos considerar comunes en toda encuesta:

Identificación del Problema. La necesidad de información para un nuevo proyecto, usualmente se presenta en forma poco clara y precisa, hay necesidad entonces de delimitar los alcances del estudio.

Formulación de Objetivos. Una vez delimitado el alcance de la encuesta se pueden concretar los objetivos e hipótesis a probar.

Conceptualización. La definición operacional de las variables involucradas, así como la escala de medición empleada surgen del marco teórico del proyecto.

Diseño de Tabulaciones. Los criterios de clasificación y grado de agregación de datos son condicionantes que el diseñador de la muestra debe conocer para determinar tamaños de muestra.

Diseño del Cuestionario. Involucra una actividad más compleja de lo que aparenta pues se pueden cometer muchos errores e inducir otros más si el cuestionario no tiene una estructura adecuada. La redacción de las preguntas del cuestionario o instrumento de recolección deben ser congruentes las definiciones operacionales y con el marco teórico adoptado.

Evaluación del marco. Se debe verificar el grado de actualización y consistencia de los datos de las unidades de la población para determinar porcentajes de exclusión de elementos de la población e inclusión de elementos extraños.

Diseño de Muestra. Resume los procedimientos de selección y estimación asociados a márgenes de errores con probabilidades conocidas.

Prueba Piloto. Es frecuente la aplicación de una prueba piloto para verificar si los supuestos planteados en los estudios de gabinete se cumplen adecuadamente. La prueba piloto permite corregir cuestionario, estimar varianzas, ajustar presupuestos y precisar tiempos de realización de trabajo de campo.

Cartografía de Enumeración. Las cargas de trabajo, requerimientos de viáticos y transportes se pueden planear eficazmente si se dispone de una buena cartografía. La cartografía es esencial para marcos definidos por áreas.

Edición de Documentos de trabajo. Manuales, instructivos y formatos de control permiten homogeneizar criterios de solución de las diversas situaciones problemáticas que se presentan en la logística de una encuesta.

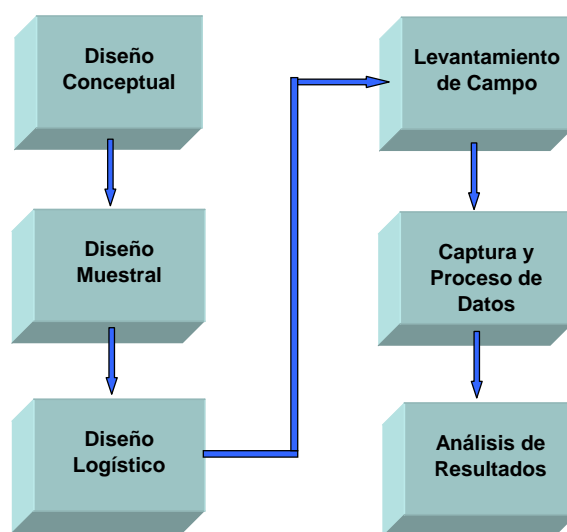
Capacitación de personal de campo. Ningún diseño de muestra por bueno que sea puede agregar calidad a datos mal registrados. La capacitación cuidadosa del personal de campo es garantía de eliminación de fuentes de error ajenas al muestreo.

Controles de campo y programación de captura. Es necesaria una excelente comunicación entre el personal de las áreas de muestreo, control de documentos en campo y los analistas y programadores responsables del proceso de datos para obtener los resultados que se esperan de una encuesta.

Codificación y captura de datos. La disposición o construcción de catálogos adecuados para una correcta codificación y captura- verificación de datos darán como resultado archivos confiables para procesos posteriores.

Proceso Estadístico de datos. La obtención de resultados mediante software específico y paquetería facilitan esta labor, pero los tabulados y reportes requieren de edición adicional para su presentación definitiva.

Edición de Informe y Anexos. Un reporte que analice resultados e incluya una memoria técnica y anexos con cuestionarios y otros documentos permitirán al usuario explotar al máximo la información.



1.8 El Cuestionario.

Los aspectos teóricos del muestreo que se comentarán en el curso suponen que se dispone de un buen instrumento para la recolección de datos, sin embargo, toda la labor desarrollada por el equipo de diseño, dirigida a lograr un buen diseño de la muestra, un adecuado control de las operaciones de campo y estimaciones óptimas, puede verse invalidada por no disponer de un cuestionario elaborado cuidadosamente.

La elaboración del cuestionario debe responder a los objetivos planteados por el proyecto. Significa la precisión de muchos aspectos, cuya conceptualización se presentaba poco clara.

En la manufactura de un cuestionario se deben tomar en cuenta las siguientes recomendaciones:

1. **Evitar el afán de información.** Cuando se elabora un cuestionario, se suele exceder el número de preguntas necesarias para alcanzar los objetivos iniciales. El exceso de preguntas fatiga al entrevistado y lo induce a falsear sus respuestas.
2. **Redactar las preguntas en lenguaje claro para el entrevistado.** Se debe considerar la heterogeneidad de capacidades, idiosincrasias y niveles culturales de los entrevistados.
3. **Hacer las preguntas concisas,** sin que ello signifique un sacrificio de la claridad.
4. **Las preguntas deben redactarse de modo que el entrevistador no tenga que participar** en la aclaración de un concepto pues puede distorsionar su intención original.
5. **Evitar** preguntas que obliguen al entrevistado a realizar **cálculos.**
6. **Evitar preguntas que molesten al entrevistado.** En ocasiones no se puede evitar el tocar temas escabrosos para el entrevistado, en estos casos se debe redactar la pregunta con el mayor tacto posible o recurrir a mecanismos de respuesta aleatorizada.
7. **Evitar preguntas que induzcan la respuesta.** El emitir en forma implícita un juicio de valor conduce respuestas inducidas y que indudablemente falsean los resultados.
8. **Considerar las tabulaciones y sus niveles de agregación** para definir las alternativas de respuesta en preguntas de opción múltiple. Clasificaciones muy amplias darán lugar a tabulaciones con celdas vacías.
9. **Presentar formatos que faciliten posteriormente la captura** y proceso de datos, con códigos claros y bien definidos.
10. **No realizar dos preguntas en una.**

Tipos de Preguntas.

Las preguntas de un cuestionario pueden clasificarse en la forma siguiente:

- . **Cerradas dicotómicas.** Preguntas cuya respuesta suele ser SI o NO y que sirven para condicionar la aplicación de bloques del cuestionario.
- . **Cerradas de opción múltiple excluyente.** La respuesta se limita a un número determinado de respuestas que mutuamente se excluyen.
- . **Cerradas de opción múltiple no excluyente.** La respuesta se limita a un número determinado de respuestas, pero el entrevistado puede seleccionar una o más respuestas simultáneamente.
- . **Abiertas.** La respuesta textual del entrevistado se transcribe para posteriormente definir criterios que permitan su codificación y proceso.

1.9 Niveles de medición.

Los niveles de medición de las variables involucradas en una pregunta constituyen otra alternativa de clasificación:

- . **Categóricas.** Su respuesta está asociada a un criterio de clasificación de dos o más clases. Los valores de las variables categóricas funcionan como elementos de identificación y no es factible realizar operaciones aritméticas con ellos. Las variables categóricas se someten procesos estadísticos que involucran frecuencias absolutas o relativas.
- . **Ordinales.** Los valores de las escalas ordinales permiten asociar un orden, pero las diferencias entre valores consecutivos de la escala no son consistentes. Son frecuentes escalas de diferencial semántico con extremos tales como totalmente de acuerdo y totalmente en desacuerdo.
- . **Intervalares.** Los datos de escalas intervalares se pueden someter a operaciones aritméticas diversas. Tienen la particularidad de que el cero de la escala no representa ausencia de la característica, sino solamente es un punto de referencia arbitrario. El caso más ilustrativo es el de la medición de la temperatura.
- . **Razón.** En esta escala son válidas todas las operaciones aritméticas y el cero representa ausencia de la característica. Por ejemplo las medidas de distancia, superficie o volumen.