

Métodos de Muestreo para la Optimización de Diseño de Experimentos

Braulio Alexis Almaguer Crespo, Dr. Edgar Cossio Franco
CIATEQ

Resumen - El diseño de experimentos es un conjunto de técnicas activas, es decir, que se manipula el experimento para que brinde la información necesaria para su mejora. El saber acerca de diseño de experimentos, junto con otras técnicas estadísticas, en combinación con el conocimiento de la naturaleza del proceso, sitúan al experimentador como un agente perceptivo y a su vez proactivo capaz de proponer mejoras y detectar áreas de oportunidad donde una persona que ignore todos estos hechos no ve nada. El fin de las técnicas estadísticas empleadas en el diseño de experimentos es lograr que el aprendizaje y conocimiento obtenido a raíz de los resultados sea lo más eficiente posible. Una técnica estadística principal es el muestreo cuyo fin es el estudiar las relaciones existentes entre la distribución de la variable objetivo en una población determinada y la distribución de dicha variable en la muestra de estudio. El objetivo de este trabajo es el de brindar un conocimiento general acerca del diseño de experimentos, elementos de inferencia estadística y de las técnicas de muestreo empleadas en el DDE, así como ejemplos útiles de aplicaciones

Palabras Clave – *Diseño de Experimentos, Aleatorización, Muestreo, Optimización*

Introducción:

En la industria, es común el hacer experimentos o pruebas con el fin de comprobar una idea o resolver un problema; el querer saber qué material es mejor para un componente o si el diseño propuesto para el mismo cumple con los requerimientos establecidos, por mencionar algunos.

El problema aquí es, las pruebas en estas situaciones se hagan con base a la intuición o experiencia, optando por la clásica metodología de prueba y error, en lugar de crear un plan experimental adecuado que asegure una respuesta específica a las necesidades del experimentador. Sin embargo, debido al avance tecnológico e industrial estas situaciones donde se opta por la metodología de prueba y error resultan ser no tan sencillas ni eficaces, para este tipo de situaciones es mejor tomar el camino que garantice resultados con base a una sustentación estadística y matemática.

El diseño de experimentos es un conjunto de técnicas activas, es decir, que se manipula el experimento para que brinde la información necesaria para su mejora. El saber acerca de diseño de experimentos, junto con otras técnicas estadísticas, en combinación con el conocimiento de la naturaleza del proceso, sitúan al experimentador como un agente perceptivo y a su vez proactivo capaz de proponer mejoras y detectar áreas de oportunidad donde una persona que ignore todos estos hechos no ve nada.

El objetivo de las técnicas estadísticas empleadas en el diseño de experimentos es lograr que el aprendizaje y conocimiento obtenido a raíz de los resultados sea lo más eficiente posible. En este proceso, podemos definir dos polos que interactúan constantemente. El primero es el polo teórico, que incluye la teoría, modelos, hipótesis, conjeturas y supuestos, véase Figura 1.

El segundo es el polo práctico, donde abarca la realidad, hechos, fenómenos, evidencias (resultados) y datos. De tal manera, al momento de crear la hipótesis del experimento, esto representa un proceso de deducción que toma en cuenta todos estos factores antes mencionados y busca ser comparada con los datos.

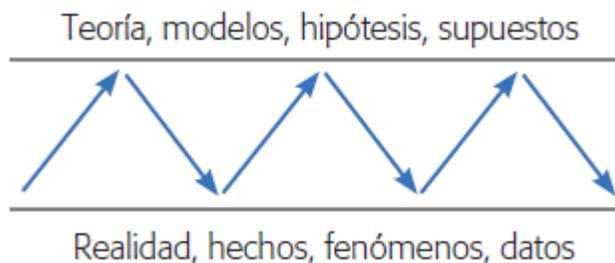


Fig. 1 Polos del Diseño de Experimentos

Cuando la hipótesis y los datos no coinciden esto deriva en un proceso de inducción, véase Figura 2, lo que significa modificar la hipótesis original, de esta manera se inicia un ciclo de iteración donde la comparación de la teoría con los datos tiene lugar, esto termina con futuras modificaciones al proceso y obtención de conocimiento.

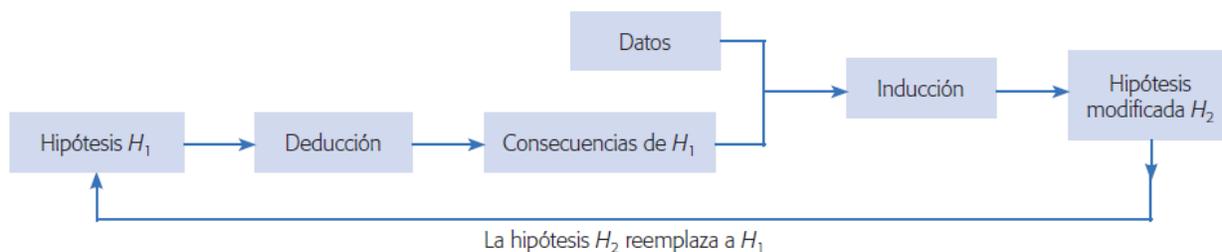


Fig. 2 Proceso de Inducción

El objetivo primordial del diseño de experimentos es estudiar la influencia de determinado tratamiento sobre un proceso, por lo que se debe experimentar aplicando el tratamiento y no aplicándolo. La metodología de diseño de experimentos analiza cómo variar las condiciones de ambiente al momento de realizar el experimento para aumentar la probabilidad de detectar algún tipo de cambio deseado y/o significativo en la respuesta, es de esta manera que se obtiene un mayor conocimiento del comportamiento del proceso en sí. La necesidad de realizar un experimento deriva principalmente de:

- Determinar las principales causas de la variación en la respuesta.
- Encontrar las condiciones ideales con las que se consigue un valor deseado en la variable de interés o respuesta.
- Comparar las respuestas en diferentes niveles de una variable de interés (variables controladas)
- Obtener un modelo estadístico-matemático del experimento en sí, para poder hacer predicciones de respuestas futuras bajo escenarios dados.

Muestreo Probabilístico

El hecho que una muestra sea representativa de la población permite al experimentador extrapolar y generalizar particularidades de la población, como previamente se ha mencionado, se pueden hacer inferencias y suposiciones de la población a partir de datos de la muestra, pero esta tiene que cumplir ciertos requerimientos para tales efectos.

Una muestra se dice que es representativa de la población si fue seleccionada al azar, es decir, que todos los sujetos o unidades dentro de la población tuvieron las mismas probabilidades de formar parte de la muestra mediante una técnica de muestreo probabilístico. Otro punto importante es que la muestra debe de poseer el tamaño número suficiente para que sea representativa con base a la distribución de la variable de estudio en la población.

Con estos requisitos cumplidos, es así como el análisis de una muestra permite realizar inferencias, extrapolaciones e inclusive conclusiones con un alto grado de certeza, se dice que la muestra es representativa de la población cuando la distribución y valor de diversas variables se pueden reproducir con márgenes de error calculables (intervalos de confianza).

Por ende, el objetivo principal del muestreo es el estudiar las relaciones existentes entre la distribución de la variable objetivo en una población determinada y la distribución de dicha variable en la muestra de estudio. Como primer paso, es fundamental delimitar los criterios de inclusión y exclusión para las unidades experimentales.

Una muestra puede ser de dos tipos: probabilística y no probabilística. Las técnicas de muestreo probabilísticas permiten al experimentador conocer la probabilidad que cada individuo del estudio tiene de ser incluido en la muestra por medio de una selección al azar. Por el otro lado, el tipo de muestra no probabilística, la selección de sus individuos de estudios depende de ciertos criterios y reglas que el propio experimentador considere prudentes debido a la naturaleza del experimento. Este tipo de muestras son consideradas poco válidas, confiables y además imposibles de reproducir dado que este tipo de muestreos no tienen una fundamentación probabilística, en pocas palabras, una muestra no probabilística no da certeza de ser representativa de la población.

Técnicas de Muestreo Probabilístico:

Aleatorio Simple:

Este tipo de muestreo garantiza que todos los individuos que conforman la población tengan la misma oportunidad de ser incluidos. Es decir, la selección de una unidad experimental al estudio no tiene relación con la probabilidad que esta tenga del resto de unidades en la población, véase Figura 3. El método consiste en:

- Definir la población objetivo
- Determinar el tamaño de la muestra representativa
- Seleccionar al azar hasta completar la muestra

En el Muestreo Aleatorio Simple se pueden distinguir 2 subtipos:

- Muestreo Aleatorio Simple con Reemplazo: Después de que un elemento sea seleccionado para formar parte de la muestra, este se devuelve y es reelegible para ser seleccionado una vez más
- Muestro sin Reemplazo: Utilizando este método, después de que un elemento sea seleccionado para la muestra no vuelve a ser devuelto a la población.



Fig. 3 Muestreo Aleatorio Simple

Ventajas del Muestreo Aleatorio Simple:

- Produce muestras representativas por lo que se puede usar estadística inferencial para el análisis de sus datos
- Cada muestra es independiente de otras, todas las posibles combinaciones de muestra tienen la misma oportunidad de pertenecer al experimento, por lo que elimina sesgos de selección.
- Procedimiento simple y directo, únicamente se utiliza una fórmula para determinar el tamaño de muestra.
- Los procedimientos estadísticos para el análisis de los datos y errores son más fáciles que con otros métodos de muestreo probabilístico

Desventajas del Muestro Aleatorio Simple:

- Tiende a tener errores de muestreo más grandes y menos preciso en comparación con el muestreo estratificado
- Puede no producir suficientes elementos de subgrupos en específicos, dejando una parte de la población fuera, si la población no es lo suficientemente grande.
- Se deja de lado el conocimiento previo del experimentador para dejar la selección al azar.

Aleatorio Estratificado:

Se le denominan estratos a los subgrupos de la población experimental que difieren en las características que van a ser analizadas y de los cuales se extraen los individuos para la muestra. La base de la creación de estratos es la definición de una variable a distinguir puede ser sexo, edad, nivel socioeconómico, dimensiones, productividad etc.

Por lo que, se divide a la población compuesta por una cantidad definida de unidades experimentales y se subdivide en X número de estratos dependiendo del rango de la variable de interés y, partiendo de esta división se hace un muestreo aleatorio simple de cada estrato para definir la muestra significativa, véase Figura 5. Se puede realizar de diferentes maneras:

- Asignación Proporcional: El tamaño de la muestra es proporcional al tamaño del estrato, respecto a la población total.
- Asignación Óptima: El tamaño de la muestra para cada estrato se encuentra definido por el experimentador, esto con el fin de utilizar el conocimiento previo de la naturaleza del experimento.

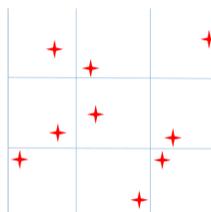


Fig. 4 Muestreo Aleatorio Estratificado

Ventajas del Muestreo Aleatorio Estratificado:

- Capacidad de estimar los parámetros de la población sino de también utilizar inferencia estadística dentro de cada estrato y comparar entre estos.
- Mejor recolección de datos de subgrupos en específico que no pueden ser capturados por el muestreo aleatorio simple
- Errores de muestreo muy pequeños en comparación al muestreo aleatorio simple
- EL muestreo aleatorio estratificado producirá AL MENOS una muestra tan precisa como una muestra aleatoria simple del mismo tamaño.
- Las muestras estratificadas tienden a ser más representación de una población debido a que se asegura que cada estrato este representado en la muestra general.

Desventajas del Muestreo Aleatorio Estratificado:

- Se necesita información sobre la proporción de la población total de cada estrato
- Es más caro y complicado que el muestreo aleatorio simple
- La selección de variables de estratificación puede ser difícil si se tienen un gran número de estas.
- Análisis más complejo que el muestreo aleatorio simple

Aleatorio Sistemático

Este tipo de muestreo tiende a ser más preciso que el aleatorio simple, puesto que se analiza a la población de una forma más uniforme. Un muestreo aleatorio sistemático se compone de los siguientes pasos:

- Se define una variable de estratificación
- De manera aleatoria, se elige un (o más) sitio de muestreo para un estrato en específico (cualquiera, aquí se puede utilizar el conocimiento del experimentador tiene sobre el experimento para elegir locaciones con mayor posibilidad de éxito)
- En los estratos restantes se utiliza la(s) misma(s) locación(es) que en el estrato principal.

Un caso típico consiste en elegir el centro de cada estrato como sitio de muestra, por ejemplo. Tanto en el muestreo sistemático como en el de estratos la forma del estrato no se encuentra sujeta a ninguna norma, esta puede ser rectangular, cuadrada, triangular etc., dependiendo de las facilidades del experimentador.

El muestreo sistemático ofrece una metodología sencilla de implementar y además brinda una cantidad suficiente se observaciones con distancias lo suficientemente considerables y direcciones bien definidas, vease Figura 5 (patrón de muestreo).

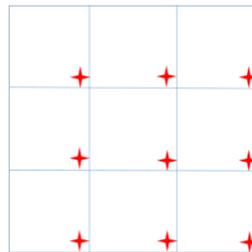


Fig. 5 Muestreo Aleatorio Sistemático

Ventajas del Muestreo Aleatorio Sistemático:

- Si el proceso de selección es manual, es una opción fácil, sencilla y que toma menos tiempo y dinero
- No se necesita enumerar la población total
- Si el orden de los elementos en el muestreo fue asignado al azar, esta opción produce resultados similares al muestreo aleatorio simple
- Elimina la posibilidad de autocorrelación de variables
- Asegura que toda la población sea considerada para la muestra

Desventajas del Muestreo Aleatorio Sistemático:

- Elementos combinados tienen menos probabilidades de ser seleccionados
- Solo la selección del primer elemento de la muestra es al azar.
- Estimación de variaciones es más complejo que en el muestreo aleatorio.

Aleatorio por Conglomerados

En el muestreo aleatorio por conglomerados se divide a la población en estratos, como es el caso del muestreo aleatorio sistemático, pero en este caso se seleccionan los estratos que van a formar parte de la muestra, puede ser de manera aleatoria o utilizando el conocimiento previo del experimentador para esta selección, Figura 7.

Una vez seleccionados los estratos, se hace un muestreo aleatorio simple hasta completar la muestra representativa. Requiere de una muestra más grande, pero suele simplificar de manera significativa las observaciones del experimento siendo más baratos y eficientes.

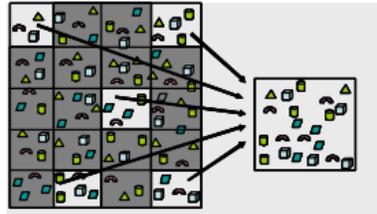


Fig. 7 Muestreo Aleatorio por Conglomerados

Ventajas del Muestro Aleatorio por Conglomerados:

- Se puede utilizar la inferencia estadística para estimar características de la población
- Se realizan varias fases de muestreo a la vez
- Eficiente en poblaciones muy grandes

Desventajas del Muestro Aleatorio por Conglomerados

- Error estándar mayor que en otros tipos de muestreo
- Cálculo del error estándar es complejo

Casos de Aplicación

El hecho de tener a disposición el conocimiento de inferencia estadística, diseño de experimentos y técnicas de muestreo brindan un conjunto de herramientas con una fuerte fundamentación matemática que puede ser de utilidad en casi cualquier industria actual. Un ejemplo de esto son los siguientes casos de aplicación, que, a través del uso correcto de técnicas de muestreo, diseño de experimentos y todo lo que esto conlleva pudieron brindar una solución eficiente, rápida y confiable

Aplicación de Técnicas de Muestreo para estimar la calidad del mosto de uvas tintas

Cambiando de tema drásticamente, otra aplicación que se estudió y resultó muy interesante fue la aplicación de técnicas de muestreo para conocer la calidad de un cultivo de uvas tintas. El principal objetivo de Llera (2017) era el de definir un criterio de muestreo probabilístico, esto con el fin de estimar los parámetros de la composición fisicoquímica de las uvas.

Esto representa un problema mayor ya que como él menciona, el momento adecuado para cosechar las uvas es cuando se tiene un equilibrio entre el dulzor de las bayas y el pH de estas, el cortar antes o después arruinaría la cosecha y la única forma de saber a ciencia cierta la calidad de toda la parcela es el hacer un recorrido y muestrear en su totalidad la parcela, una opción demorada y costosa.

En este trabajo explica su metodología para conocer con una estimación del 95% la calidad del mosto de una cosecha de no una, sino 4 parcelas a cosechar. Se aplicaron 3 muestreos sistemáticos consecutivos a 1 parcela, posteriormente a 10 racimos y finalmente a 100 bayas.

En cada caso de técnica empleada se conoció con una estimación de intervalos el contenido azucarino, pH y acidez total del mosto. Indica que la mayor exactitud de resultados se obtuvo con muestreo sistemático, puesto que como variables de control utilizó las parcelas y posteriormente subdividió en racimos y posteriormente en bayas, es decir un muestreo sistemático anidado.

Como conclusión, menciona que una muestra representativa de la parcela solo puede ser obtenida utilizando técnicas de muestreo sistemático con arranque aleatorio y también tomando en consideración a todo el viñedo, sin excluir alguna parte de la población, confirmando una vez más lo anteriormente mencionado relacionado a la inferencia estadística.

Diplomado en Análisis de Información Geoespacial: Tipos de Muestreo

En este trabajo extraído de la página del Centro Público de Investigación CONACYT, el Mtro. Alberto Porras (-), nos explica los diferentes tipos de muestreo que se emplean en los estudios de población geográfica como

lo son una encuesta o un censo. Menciona que el primer paso de esta clase de estudios es el de generar una muestra representativa, que coincide con los otros dos casos anteriormente expuestos, pero que como se trata de un estudio de índole geoespacial también se deben considerar espacios geográficos, como lo son el mar y la altura, por lo que recomienda que en esta clase de estudios se realice en el mismo lugar, varias muestras, pero a diferentes instancias de tiempo.

Tocando el tema de los factores de elección, como lo es el mar y altura en este trabajo, nos indica también que, para esta clase de estudios, la muestra debe estar conformada por una población lo suficientemente grande como para ser representativa, pero también, dichos elementos deben ser representativos del fenómeno que se busca explicar, por lo que las unidades también deben de ser seleccionadas con base a criterios de unidades, elementos, composición, áreas y periodos de tiempo.

Porras también hace la sugerencia de empezar el muestreo con una delimitación geográfica y, partir a términos más específicos, como es el caso de ciudad, después colonia y posteriormente calle y, que una vez se tenga el término geográfico mínimo, partir con técnicas de muestreo.

Conclusiones

El diseño de experimentos es un conjunto de técnicas activas que, acompañado del conocimiento previo del experimentador, ayudan a brindar conocimiento seguro, rápido y eficiente al menor costo. Se compone de una serie de pasos ya previamente definidos, pero, como se ha expuesto previamente, varían enormemente dependiendo de la naturaleza del fenómeno que se pretende explicar.

Sin embargo, independientemente de la variable a analizar, el tipo de estudio que se está realizando o el dimensionamiento de la población (como pudimos ver en los casos de aplicación) el tener una muestra representativa de la misma es una pieza clave para el éxito del experimento. El uso de la inferencia estadística nos ha brindado un conjunto de herramientas que ayudan a potencializar nuestros procesos, cualesquiera, y el uso correcto de las técnicas de muestreo resultan ser primordiales en el diseño de experimentos.

Referencias:

- Casal, J., et al (2003) *Tipos de Muestreo*, Universidad Autónoma de Barcelona, Recuperado de [http://mat.uson.mx/~ftapia/Lecturas%20Adicionales%20\(C%3%B3mo%20dise%C3%B1ar%20una%20encuesta\)/TiposMuestreo1.pdf](http://mat.uson.mx/~ftapia/Lecturas%20Adicionales%20(C%3%B3mo%20dise%C3%B1ar%20una%20encuesta)/TiposMuestreo1.pdf)
- Gutierrez, H. (2008), *Análisis y Diseño de Experimentos*, México D.F.: McGrawHill
- Hernandez, C.J. et al., (2019) *Estadística y Probabilidad I*, Universidad Nacional Autónoma de México, Recuperado de: https://cch.unam.mx/sites/default/files/recursos_files/PAE_Estadistica1.pdf
- Kleeberg, F., et al (2009) *Aplicación de las técnicas de Muestreo en los negocios y en la industria*, Universidad de Lima, Recuperado de: <https://www.redalyc.org/pdf/3374/337428493002.pdf>
- Llera, J., et al (2015) *Aplicación de Técnicas de Muestreo Probabilístico para estimar la calidad del mosto de uvas tintas*, Universidad Nacional de Cuyo Argentina, Recuperado de: <https://www.redalyc.org/pdf/3828/382852189010.pdf>
- Marchant, J.I.(2016), *Optimización de Muestreo Espacial en Experimentos de Difusión en Materiales*, Tesis de Licenciatura, Universidad Técnica Federico Santa María, Recuperado de : <https://repositorio.usm.cl/bitstream/handle/11673/23224/3560900232317UTFSM.pdf?sequence=1&isAllowed=y>
- Marin, J.M. (2012), *Introducción al Diseño de Experimentos*, Recuperado de: <http://halweb.uc3m.es/esp/Personal/personas/jmmarin/esp/Diseno/IntroDE.pdf>
- Salinas, A.M. (2004) *Métodos de Muestreo*, Departamento de Ciencias de la Universidad Autónoma de Nuevo Leon, Recuperado de: <https://www.redalyc.org/pdf/402/40270120.pdf>
- Porras, A. (-) *Diplomado en Análisis de Información Geoespacial: Tipos de Muestreo*, Centro de Investigación en Geografía y Geomática “Ing. Jorge L. Tamayo” Recuperado de: <https://centrogeo.repositorioinstitucional.mx/jspui/bitstream/1012/163/1/19-Tipos%20de%20Muestreo%20-%20%20Diplomado%20en%20An%C3%A1lisis%20de%20Informaci%C3%B3n%20Geoespacial.pdf>
- Portela, J., et al (2007) *Técnicas Básicas de Muestreo con SAS*, Repositorio de la Universidad Complutense de Madrid, Recuperado de: <https://eprints.ucm.es/id/eprint/47107/2/T%C3%A9cnicas%20b%C3%A1sicas%20de%20muestreo%20con%20SAS.%20J.%20Portela.%20M.%20Villete.pdf>