

Aspectos estadísticos, técnicos y sociológicos del Diseño de Experimentos

Ing. Daniel Firka

Druida Software & Consultoría (www.druida.biz)

Instituto Argentino para la Calidad (www.iapc.org.ar)

El diseño experimental (comúnmente llamado DOE por sus siglas en inglés) es una herramienta muy poderosa para comprender y mejorar procesos productivos. Sin embargo y lamentablemente, su uso es bastante limitado en la práctica cotidiana de las organizaciones.

El diseño de experimentos es el estudio a través de uno o más ensayos del efecto de múltiples variables de entrada (X's) sobre una o más variables de respuesta (Y's) en un proceso determinado. Usaremos el término "DOE" para referirnos al proceso que comienza en un problema específico y continúa con el diseño, la experimentación y finalmente el análisis de los resultados.

En este trabajo se expondrán una serie de limitaciones y problemas de diversa índole que dificultan la aplicabilidad de esta herramienta en el contexto de organizaciones reales.

El análisis se hará diferenciando tres dimensiones:

- Dimensión Estadística: relacionada con los supuestos estadísticos y los métodos matemáticos que brindan validez a las técnicas DOE.
- Dimensión Técnica/Metodológica: relacionada con la forma de llevar a cabo los experimentos para sacar el máximo provecho al DOE
- Dimensión Sociológica/Gerencial: relacionada con la organización y administración del uso de la herramienta en el contexto organizativo.

Las recomendaciones se nutren de las siguientes fuentes primarias y secundarias:

- La experiencia propia del autor en experimentos industriales actuando como consultor o entrenador.
- La bibliografía existente relacionada con el diseño experimental
- Trabajos específicos publicados en revistas especializadas.

Dimensión Estadística

A pesar de la existencia de numerosos paquetes informáticos que permiten evitar errores en los cálculos, existen otros problemas que pueden afectar negativamente los análisis en DOE.

No identificar los riesgos alfa y beta en la etapa de diseño para calcular el número de réplicas

Es importante identificar el número de réplicas que se realizarán teniendo en cuenta la precisión buscada en términos del problema práctico. En el contexto de DOE, el riesgo alfa se relaciona con la

probabilidad de una falsa alarma, es decir identificar un factor como importante cuando no lo es. Por otro lado, el riesgo beta se refiere a perder señales, no detectando factores relevantes.

Definir los riesgos alfa y beta de antemano permite calcular este número de réplicas y define la sensibilidad del diseño, si se realizan muy pocas corridas experimentales, habrá mucha incertidumbre al identificar factores importantes, por otro lado excesivas corridas no sólo incurren en un costo muy alto en tiempo y dinero, sino también hacen el experimento demasiado sensible a factores poco relevantes.

Un error común es creer que cuantos más datos se incorporan en el análisis, mejor. Por ejemplo al comparar dos niveles de un factor (como velocidad de la máquina) cuyo efecto es muy leve, el analista toma datos de un PLC para todo un mes de producción en ambos factores, haciendo un test t con 1000 unidades en cada condición, resultando una diferencia estadísticamente significativa. El siguiente error del analista es pensar “este factor es importante porque resultó significativo”.

Tratar variables que representan porcentajes defectuosos como si fueran variables numéricas continuas.

Variables que representan el número de unidades defectuosas (generalmente modeladas con la distribución binomial) o número de defectos (modeladas con la distribución de Poisson) pueden ser estudiadas como variables de respuesta, pero debe tenerse en cuenta esta situación a la hora de calcular el número de réplicas, y también verificar que no se requiera una transformación una vez realizado el experimento.

Para fórmulas sobre como calcular el número de replicas en estas situaciones, ver Diamond (1989)

Tratar un diseño que debería ser “por parcelas” (Split-plot) como un diseño factorial corriente.

Cuando alguno de los factores es muy difícil de cambiar, es recomendable analizar el diseño como un diseño por parcelas, en vez de hacer un full factorial.

De la misma manera, en ocasiones los factores están “anidados”, por ejemplo si queremos comparar la vida útil de un producto que puede tener dos formulaciones posibles, que en el proceso tienen que pasar por un horno durante un cierto tiempo, y este horno puede estar en dos temperaturas, decimos que el factor formulación está anidado con el factor horno.

No tener en cuenta el problema de comparaciones múltiples de efectos

En muchas situaciones, el experimento analiza simultáneamente una serie de factores, y luego se realizan test de hipótesis para evaluar la significancia estadística de los mismos. Al hacer una serie de test de hipótesis, cada uno con cierto nivel de significación alfa, la probabilidad de que al menos uno de ellos se verifique significativo es mayor que $\alpha 1 - (1 - \alpha)^k$ (k es el número de factores involucrados). Esto significa que es más probable identificar erróneamente ciertos factores como significativos.

Esto debe ser tenido en cuenta y brinda racionalidad a la necesidad de efectuar experimentos confirmatorios donde exigimos un p-value inferior para poder confirmar el efecto sospechado en uno o más factores.

Ausencia de análisis de patrones en los residuos

Los residuos representan la variación no explicada mediante nuestro modelo experimental. En los modelos más usados, el criterio para identificar factores relevantes exige que los residuos tengan un cierto comportamiento (distribución normal, independencia y homogeneidad de su variabilidad). Con el interés en identificar los factores más importantes es fácil olvidar la verificación del comportamiento de los residuos, lo que puede invalidar las conclusiones sobre significatividad estadística. Otro problema común es considerar exclusivamente la normalidad de los residuos, cuando los aspectos más importantes del análisis de residuos son su independencia, la homogeneidad de varianzas y la normalidad, en ese orden.

No identificar datos muy alejados (*outliers*)

Los datos muy alejados o *outliers* son una fuente muy rica de información sobre el proceso, generalmente representan “causas especiales” señalando oportunidades para profundizar nuestro conocimiento.

Cuando el analista se encuentra con un dato muy alejado, la reacción no debe ser de frustración ante una corrida experimental perdida, sino de curiosidad y expectativa para hallar la causa raíz de la discrepancia.

Aplicar transformaciones a los datos sólo “porque es fácil”

Hay dos objetivos inmediatos de aplicar una transformación: estabilizar la varianza o hacer la distribución más simétrica. Por ejemplo: aplicar la raíz cuadrada a una variable consistente con la distribución de Poisson logra ambos objetivos, así como aplicar el arco seno a una variable binomial.

Sin embargo las transformaciones no son inocuas y muchas veces ocultan más de lo que permiten descubrir al analista, haciendo al mismo tiempo difícil comunicar los resultados del experimento.

El peor uso de la transformación es cuando el analista busca eliminar la ocurrencia de un outlier a través de una transformación de la variable; esto representa el triunfo de las maniobras matemáticas sobre el sentido común.

Olvidar interacciones confundidas con efectos principales

Es importante tener en cuenta que las interacciones entre dos factores ocurren frecuentemente en la práctica. Cuando usamos un diseño experimental de cribado donde buscamos identificar factores relevantes sobre un grupo muy grande de factores candidatos, muchas veces los efectos de los factores se hayan confundidos con interacciones de dos o más factores, y es posible concluir que un factor es importante aunque en realidad lo importante es la interacción confundida.

Extrapolar resultados fuera del área de operación

Cuando se investiga la superficie de respuesta en función de factores causales, es tentador buscar seteos óptimos de operación en valores de los factores muy alejados del rango que se utilizó para los factores durante el experimento. Esto puede llevar a predicciones inválidas o con una incertidumbre muy alta.

Cuando el analista sospecha que el área donde se ubica el óptimo esta fuera de los niveles utilizados en el experimento es recomendable que desarrolle otro experimento.

Dimensión Técnica / Metodológica

Tomar datos históricos (estudio observacional) y procesarlos como si fueran resultado de un experimento (estudio experimental)

La facilidad con que los modernos paquetes informáticos pueden digerir información y analizar cualquier set de datos puede tentar al analista a procesar incorrectamente un set de datos histórico como si se tratara del resultado de un experimento. Esto generalmente lleva a conclusiones erróneas, dado que la falta de control involucra mucha dispersión no explicada y fluctuaciones del proceso que luego incrementan la magnitud del ruido subyacente, disminuyendo el poder de la herramienta para identificar los factores de alto impacto.

Muchas veces los estudios observacionales son puramente exploratorios y se bucea en los datos tratando de identificar factores de interés. A veces en estos casos se trabaja con p-values más altos para descubrir el mayor número de variables sospechosas. Esto es válido siempre y cuando se entiendan los riesgos de trabajar con datos del proceso en estado “salvaje”. Puede haber variables no medidas que generan efectos en variables que sí medimos y que incorrectamente confundimos como causales. Puede encontrarse alta autocorrelación en la serie de datos, lo que afecta la validez de los test de hipótesis, etc.

Un error típico sobre todo para aquellos que toman el primer curso de DOE y vuelven a aplicar inmediatamente la herramienta en su organización es utilizar el mismo set de datos históricos para seleccionar variables para el modelo y para determinar su significación. Siempre se debe seguir un proceso secuencial donde la confirmación es posterior a la selección (Good, 2002)

No asegurar la estabilidad del proceso en el momento de realizar el experimento.

Hay discrepancia en la bibliografía sobre la cuestión de control estadístico y diseño experimental. Algunos autores recomiendan asegurar la estabilidad del proceso antes de comenzar a experimentar; otros arguyen que la aleatorización y el bloqueo son técnicas que permiten ignorar el estado de control estadístico como precondition para experimentar (para una discusión entre varios autores sobre este tema, ver Bisgaard, 2008)

Exigir controlar estadísticamente el proceso antes de comenzar a experimentar puede hacer demorar mucho el inicio de las corridas. Por otro lado, experimentar sobre un proceso totalmente impredecible genera un ruido excesivo que dificulta la identificación de factores significativos. Un camino medio es recomendable: no esperar hasta que el proceso este absolutamente bajo control, pero tampoco trabajar sobre un proceso caótico, y durante el experimento estar atentos a la aparición de causas especiales.

No atravesar etapas rigurosas de Planificación y Diseño, pasar directamente a la experimentación

Muchas veces se olvida definir claramente el problema que se quiere atacar; dado que esta tarea previa requiere la coordinación de esfuerzos entre todos los involucrados en el problema, operativamente

resulta más fácil definir los factores e informar que se va a realizar el experimento, que pasar por las etapas previas de definir la variable de respuesta, los factores y sus niveles y el tipo de problema a resolver. Este riesgo de aislacionismo es mayor cuando el practicante no tiene recursos de coaching o acompañamiento disponible y está presionado a mostrar resultados concretos en cortos plazos.

No involucrar jugadores claves para identificar variables

Toda vez que se plantee un diseño experimental es fundamental involucrar todos aquellos actores que pueden aportar conocimiento para:

- Definir claramente la variable de respuesta
- Identificar todos los factores involucrados, incluso aquellos que a primera vista pueden significar poco relacionados.
- Ayudar a definir realísticamente los niveles de cada factor.

Olvidar bloquear por alguna condición importante.

El bloqueo es uno de los tres pilares de la experimentación (junto a la aleatorización y la replicación), y permite aumentar la precisión del experimento cuando conocemos de antemano ciertas fuentes de variabilidad. Por ejemplo, si sospechamos que entre dos días consecutivos pueden actuar variables que afecten nuestro proceso y no podemos realizar todas las corridas el mismo día, es importante considerar el “día” como un bloque de manera que la variación entre días no quite precisión a nuestro experimento. En general los bloques corresponden a factores de ruido que no me interesan y cuya variabilidad quiero eliminar en mi análisis.

No asegurar que las variables no incluidas en el experimento se mantienen controladas durante la ejecución del experimento

La ejecución del experimento puede considerarse una “mesa de operaciones” donde –a excepción de los factores que estamos estudiando- se debe buscar un estado “ceteris paribus”, que es artificial del punto de vista práctico pero es funcional a nuestro objetivo de identificar el efecto de factores.

Es recomendable que el analista esté presente durante la ejecución de las corridas experimentales, y preste especial atención al control de las condiciones ambientales.

Usar condiciones o unidades experimentales que no son similares a lo que ocurre en la práctica

Esto puede suceder cuando se planea un experimento que requiere una preparación especial; en estos casos a veces es necesario solicitar material de prueba a proveedores, o preparar probetas para realizar los experimentos, y se corre el riesgo de, voluntaria o involuntariamente, trabajar sobre material o condiciones alejadas de lo que ocurre en el proceso en la realidad.

En otros casos existe una planta piloto donde se experimenta, y los resultados luego son incorporados al proceso real. Es importante estudiar la correlación entre la planta piloto y el proceso, dado que puede haber características físicas o químicas que impiden extrapolar ciegamente entre ambos entornos.

No verificar la validez de los instrumentos de medición utilizados tanto para las variables de respuesta como para los factores

Todas las observaciones son resultado de sistemas de medición que agregan una fuente de variabilidad a los datos. Si esta variabilidad adicional es muy alta impedirá medir el efecto de los factores sobre la variable de respuesta.

En ciertas ocasiones resulta incluso recomendable utilizar instrumentos de mayor precisión durante la experimentación que los utilizados rutinariamente en el proceso.

Elegir niveles inadecuados de los factores estudiados

Los niveles de los factores o categorías a considerar se deben elegir cuidando de que estén suficientemente separados para discriminar y medir su efecto pero no tanto para ser irreales o imprácticos.

En diseños de factores en dos niveles elegir los niveles muy alejados puede esconder efectos de curvatura (esto puede prevenirse incorporando puntos centrales para verificar la ausencia de curvatura)

No estratificar en casos de productos o procesos disímiles (querer resolver el hambre del mundo con un DOE)

Cuando se tiene un presupuesto acotado en recursos económicos, temporales o humanos, es riesgoso intentar cubrir un amplio abanico de problemas con un solo set de experimentos. Cada proceso quizás exige su propio estudio, y en vez de incorporar los distintos tipos de productos como niveles de un factor "producto", es muchas veces mejor trabajar sobre un producto determinado y luego una vez comprendido este proceso intentar extrapolar y avanzar a otros productos similares.

Mantener niveles de un factor constantes entre replicas consecutivas.

Es frecuente cuando se aleatorizan las corridas experimentales, que haya dos o más corridas donde un factor se encuentra en el mismo nivel. Supongamos un experimento donde uno de los factores es la temperatura del horno que puede estar en nivel alto (140°) y nivel bajo (120°), y al preparar la lista de corridas experimentales, obtenemos un patrón como el siguiente:

Corrida	Factor 1	Factor 2	...	Temperatura	...	Factor N
...						
6				120		
7				120		
8				140		
...						

Cuando el cambio de temperatura en el horno es un procedimiento trabajoso, existe la tentación de dejar el horno a 120° entre la corrida 6 y la 7. Sin embargo, es recomendable "resetear" los factores y permitir que el horno baje de temperatura y vuelva a subir al valor del nivel correspondiente, para incorporar en el sistema la variabilidad correspondiente a situaciones lo más parecidas posibles al comportamiento real del proceso.

Problemas en la toma de datos

Cuidado con descuidar la operatoria de toma y volcado de los datos para su posterior análisis. Wu y Hamada recomiendan tener hojas separadas para cada combinación de factores, para evitar confusión que pueden generar resultados incorrectamente asignados a una combinación determinada.

Dimensión Sociológica/Gerencial

Gastar excesivos recursos en el primer set de experimentos.

El entusiasmo inicial en el uso de DOE puede llevar a planificar experimentos con gran cantidad de factores y niveles. En los estadios tempranos del uso de DOE se tiene poca práctica en la identificación y determinación de niveles de factores, esta situación es problemática porque:

- la falta de experiencia puede generar problemas en la forma de planificar, llevar a cabo o interpretar los resultados.
- DOE en un contexto industrial es siempre un proceso iterativo y secuencial, donde cada sucesivo experimento va disparando nuevas preguntas y respondiendo otras, hasta que el equipo determina que el conocimiento adquirido permite alcanzar un nivel aceptable de excelencia en el proceso involucrado.

Aplicar DOE por razones ajenas a la mejora del proceso o producto

La currícula estándar de black-belts en la metodología Seis Sigma tiene generalmente una descripción de métodos de diseño experimental, que en muchos casos se da fugazmente y en conjunto con muchas otras herramientas en el marco de la semana 2 o 3 de entrenamiento, para ser utilizado en las fases de análisis o mejora del esquema DMAIC. Esta situación a pesar de ser beneficiosa por enseñar la herramienta, conlleva el peligro de intentar aplicarlo sólo para atravesar satisfactoriamente los “peajes” establecidos para el proyecto de mejora.

En otros casos la motivación para el DOE es jugar con el software recientemente adquirido. El diseño experimental y el software asociado siempre es un medio para lograr el fin de optimizar procesos, nunca un fin en sí mismo.

Una persona del área de Investigación y Desarrollo puede recibir una capacitación intensiva en DOE y luego tener una presión para mostrar rápidamente los resultados prácticos. El objetivo se convierte en demostrar lo aprendido más que optimizar un proceso.

No sensibilizar a los mandos gerenciales en la utilidad y ventajas que puede reportar el uso adecuado de DOE.

Este es un factor crítico para la sustentabilidad del diseño experimental. Si los niveles gerenciales ven el DOE como una herramienta compleja para unos “elegidos”, sobre la que no tienen control ni conocimiento, fácilmente podrán disminuir el valor simbólico asociado a la herramienta en su esquema mental. Esto puede prevenir la locación de recursos necesarios para llevar a cabo experimentos eficaces, que en muchos casos requieren de un proceso largo e iterativo.

No sensibilizar a los operadores y otros actores relacionados con el proceso en la utilidad y ventajas que puede reportar el uso adecuado de DOE.

Así como los niveles directivos deben conocer la herramienta para poder habilitar recursos, los operarios y mandos de línea tienen que estar al tanto para obtener su apoyo a la hora de buscar factores, niveles e implementar mejoras.

Utilizar una experiencia poco exitosa como “excusa” para no volver a planear experimentos

Muchos de los factores expuestos previamente pueden hacer que los primeros experimentos resulten fallidos. Estas fallas iniciales pueden catalizar una sensación negativa, generando un círculo vicioso entre “no utilizamos DOE porque no funciona” y “no funciona porque no le damos la oportunidad para mostrar resultados”.

No comunicar claramente los resultados del experimento a las partes interesadas.

Siempre tener en cuenta que parte de la audiencia que leerá las conclusiones del experimento no tiene el nivel de conocimientos para interpretar estadísticamente los datos. Esto exige basarse principalmente en gráficos de efectos e interacciones, explicando las razones por las cuales se llegaron a las conclusiones minimizando la jerga estadística y concentrándose en la relevancia práctica y la claridad conceptual.

Conclusiones

Toda compañía comenzando el uso del diseño experimental debe ser consciente de los problemas potenciales y sugerencias para aprovechar al máximo esta poderosa herramienta.

El presente trabajo no pretende reemplazar la abundante bibliografía sobre el tema, sino destacar algunos aspectos claves que los practicantes deben tener en cuenta a la hora de embarcarse en el uso de DOE.

Bibliografía

Bisgaard, S (2008): "Must a Process be in Statistical Control before Conducting Designed Experiments?", *Quality Engineering*, ASQ, pp 143 - 176, Vol 20, Nr2

Box, G., Hunter & Hunter (2005): *Statistics for Experimenters: Design, Innovation, and Discovery*. (Wiley-Interscience; 2 edition. USA)

Corwin, AI (1999): "How to Institute DOE in Your Company" available at http://www.strategy4doe.com/Newsletters/BackIssues/Vol2/September_1999.htm

Costa, N (2006): "Guidelines to help practitioners of design of experiments". *The TQM Magazine*. Volume 18 Issue 4 pp. 386 – 399

Diamond (1981): *Practical Experimental Design*. (Lifetime learning Publications. USA)

Good, P. and Hardin, J. (2003): *Common Errors in Statistics and how to avoid them*. (Wiley. USA).

Hybarger, Jeff (2007): "The ten most common designed experiment errors" available at <http://www.dpnCanada.com/images/stories/2007/June/PDFs/dpn%20june%2007-26.pdf>

Ilzarbe, L (2008): "Practical applications of design of experiments in the field of engineering: a bibliographical review". *Quality and Reliability Engineering International*. Volume 24 Issue 4, pp 417 – 428

Kowalski, S. & Potcner, K (2004): "How to Analyze a Split-Plot Experiment" *Quality Progress*, Vol. 37, No. 12, December 2004, pp. 67-74

Lindsey and Lambert (1998): "On the appropriateness of marginal Models For Repeated Measurements in Clinical Trials", *Statistics in Medicine* 17, pp. 447-469

Montgomery (2008): *Design and Analysis of Experiments*. (Wiley. USA)

Novig, Peter (undated): "Warning Signs in Experimental Design and Interpretation" available at <http://norvig.com/experiment-design.html>

Tanco, E et al (2008): "How is experimentation carried out by companies? A survey of three European regions". *Quality and Reliability Engineering International*. Volume 24 Issue 8, pp 973 – 981

Van Belle, G (2002): *Statistical Rules of Thumb*. (Wiley Inter Science. USA)

Wheeler, Donald J. (1990). *Practical Experimental Design*. (SPC Press. Tennessee, USA)

Wheeler, Donald J. (2004). *The Six Sigma Practitioner's Guide to Data Analysis*. (SPC Press. Tennessee, USA)

Wu, J, Hamada, M (2000): *Experiments: Planning, Analysis, and Parameter Design Optimization*. (Wiley-Interscience. USA)