Aplicación de un Diseño de Experimentos 2^k para la Optimización de un Proceso de Moldeo por Inyección

Gloria Paola Ortiz Espinoza,Carlos Anaya Eredias, Gilberto Ortiz Suárez, Pedro David Sánchez Pérez

Universidad de Sonora, Departamento de Ingeniería Industrial, Rosales y Blvd. Luis Encinas S/N, C. P. 83000, Hermosillo, Sonora, México.
gloriapa.ortizes@correoa.uson.mx, canaya@industrial.uson.mx, gortiz@industrial.uson.mx, pedroda.sanchezpe@correoa.uson.mx

Resumen. Tener variabilidad en el proceso de producción es inevitable debido a inconsistencia en los parámetros de la máquina, la pieza de trabajo, materiales, herramientas y procesos. Cuando existe variación en el proceso el diseño de experimentos es una técnica estadística utilizada para determinar el ajuste óptimo de los factores que influyen al proceso, y de esta manera mejorar su funcionamiento, reducir su variabilidad y mejorar la manufactura de productos. El presente artículo se enfoca en la identificación, análisis y estudio de los factores principales que influyen durante el proceso de moldeo por inyección de la pieza 1-1532082-9 para que el producto cumpla con los requerimientos de calidad exigidos por el cliente.

Palabras Clave: Proceso, Moldeo por inyección, Diseño de experimentos, Factores

1 Introducción

Dentro de la empresa TE Connectivity se lleva a cabo el proceso de moldeo por inyección de la pieza 1-1532082-9, donde se observa que los diámetros superiores se encuentran fuera deespecificación, es decir, no están dentro del rango de medida de 0.118 a 0.122 mm. Cuando la pieza no cumple con dicha especificación de calidad, ésta debe de ser retrabajada, sin embargo en ocasiones resulta imposible, y pasa a formar parte del desperdicio, generando costos de calidad para la empresa.

Este estudio tiene como objetivo identificar, analizar y estudiar los factores principales que influyen durante el proceso de moldeo por inyección de la pieza, para para que el producto cumpla con los requerimientos de calidad exigidos por el cliente. La estructura

del artículo comienza con el marco teórico que proporciona el sustento necesario y suficiente para el desarrollo de un diseño de experimentos en un proceso de moldeo por inyección. Así mismo se explica la metodología a utilizar y se presentan el análisis de los datos. Para finalizar se presentan los resultados obtenidos durante el proyecto y se desarrollan las conclusiones del artículo.

2 Marco Teórico

La calidad de los productos y servicios se ha convertido en uno de los factores principales del funcionamiento óptimo de una organización, por lo que los fabricantes han tomado una creciente conciencia de la necesidad del mejoramiento continuo de sus procesos para obtener y mantener buenos resultados económicos en el desempeño de sus organizaciones [1].

El moldeo por inyección representa el proceso más importante para manufacturar partes plásticas debido a la facilidad de fabricación de artículos de alta calidad [2], además tiene algunas ventajas como una alta estabilidad dimensional, ciclos cortos de producción, superficies finas y limpias en los productos, moldeo fácil de formas complicadas y bajos costos en la producción [3], es por esto que para obtener una alta calidad en piezas moldeadas por inyección, el diseño de moldes, la materia prima, condiciones de procesamiento y el rendimiento de la máquina de inyección debe estar bien controladas para cumplir con las especificaciones requeridas.

2.1 Problemas en el Moldeo por Inyección

Los defectos de los productos, tales como deformaciones, encogimiento, marcas de hundimiento, y tensión residual son causados por diversos factores durante el proceso de producción. Estos defectos influyen en la calidad y la precisión de los productos, por lo tanto, es de suma importancia controlar eficazmente la influencia de los factores durante el proceso de moldeo [4]. La complejidad del proceso de moldeo por inyección y la enorme cantidad de manipulación de parámetros en tiempo real crean un esfuerzo muy intenso para mantener el proceso bajo control. Lo que es más, la complejidad y la manipulación de los parámetros pueden causar graves problemas de calidad y altos costos de fabricación [5].

El moldeo por inyección es un proceso cíclico inestable, sin embargo, la selección apropiada de los factores para su correcta operación, se hace más difícil conforme el diseño de la pieza de plástico se hace más delgada [6]. Debido a la complejidad del moldeo por inyección, numerosas propuestas y modelos matemáticos han sido ampliamente desarrollados ya que se han realizado numerosos estudios para el análisis de las diferentes etapas del proceso de moldeo por inyección [6].

2.2 Diseño de Experimentos

Tener variabilidad en el proceso de producción es inevitable debido a inconsistencia en los parámetros de la máquina, la pieza de trabajo, materiales, herramientas y procesos [7]. Cuando existe variación en el proceso el diseño de experimentos es una técnica estadística utilizada para determinar el ajuste óptimo de los factores que influyen al proceso, y de esta manera mejorar su funcionamiento, reducir su variabilidad y mejorar la manufactura de productos [8]. Es utilizado para entender las características del proceso y para investigar cómo las entradas afectan las respuestas basadas en antecedentes estadísticos, además de determinar los parámetros óptimos del proceso con un menor número de ensayos de prueba [9].El diseño de experimentos es altamente efectivo para aquellos procesos, que su rendimiento se ve afectado por varios factores.

2.3 Diseño Factorial 2^k

Los diseños factoriales se usan ampliamente en experimentos que incluyen varios factores cuando es necesario estudiar el efecto conjunto de los factores sobre una respuesta. El más importante de estos casos especiales es el de k factores, cada uno solo con dos niveles. Estos niveles pueden ser cuantitativos, como dos valores de temperatura, presión o tiempo, o bien cualitativos, como dos máquinas, dos operadores, los niveles "alto" y "bajo" de un factor, o quizás la presencia o ausencia de un factor. Una réplica completa de este diseño requiere 2 x 2 x … x 2 = 2^k observaciones y se le llama diseño factorial 2^k[10].

3 Descripción del Problema

El impacto negativo que se tiene cuando la pieza 1-1532082-9 no cumple con los requerimientos de calidad exigidos por el cliente se puede medir en dos formas:
1. En el departamento de moldeo los índices de desperdicio de silicón varían entre el 40% y 50% del material productivo, además, cuando la pieza puede ser retrabajada se invierte en promedio 20 horas a la semana para hacerlo.
2. El componente no se entrega a tiempo al ensamble final del número de parte 6-1532028-0 de la línea de producción Microdot, ocasionando que la orden no se pueda trabajar, incrementando de esta manera el WIP, es decir órdenes que no se han completado y están esperando su procesamiento posterior, y dejando sin embarcar 7,214 dólares americanos por cada orden, teniendo en total 17 ordenes sin trabajar acumulando 122,638 dólares americanos hasta el momento.

4   Implementación

En general la experimentación se hace de manera secuencialy antes de comenzar con el experimento se debe de cerciorar que los instrumentos de medición y la máquina con la que se va a trabajar estén dentro de periodo de calibración y mantenimiento. Esto ayudará a que la variable de respuesta sea más exacta y el error de medición disminuya. En la Figura 1 se muestra el esquema general del procedimiento para diseñar un experimento el cual comprende 5 pasos los cuales se describen a continuación:
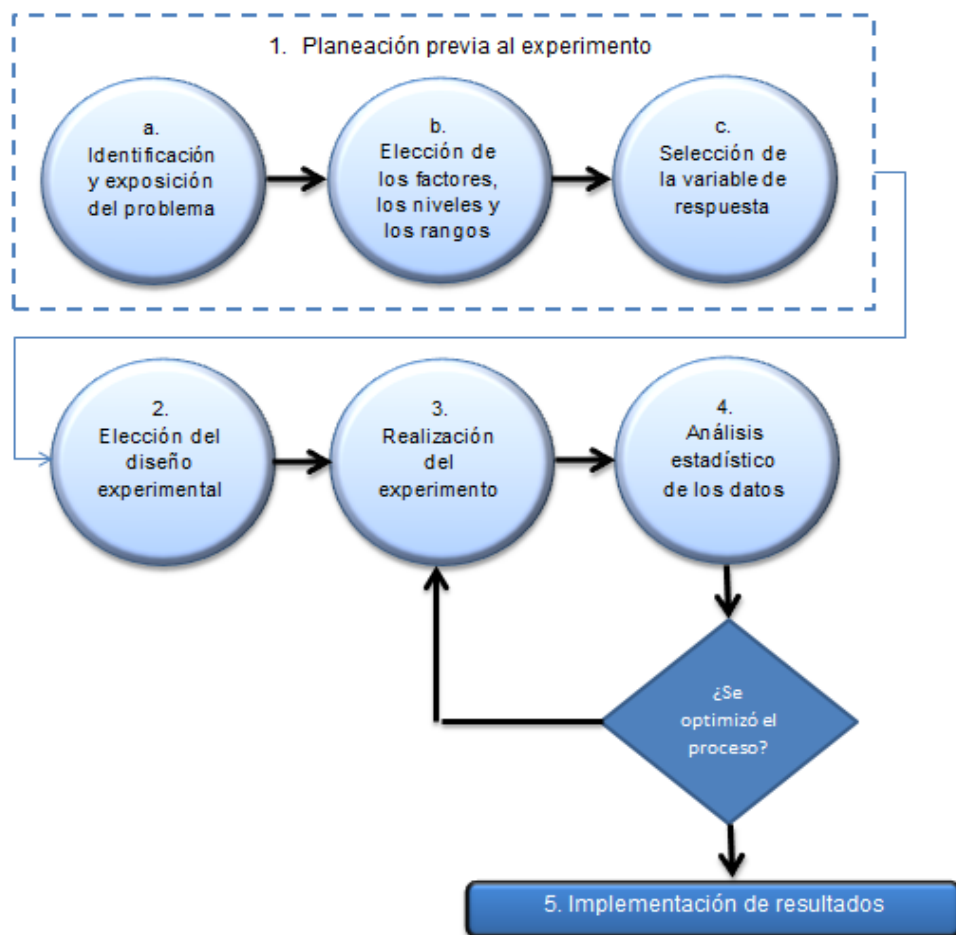
Figura 1. Metodología para realizar un diseño de experimentos

4.1   Planeación Previa al Experimento

Identificación y Exposición del Problema

Se recolectaron 90 datos de cada orificio obtenidos en 3 días de producción para los cuales se realiza un diagnóstico gráfico utilizando de las siguientes herramientas estadísticas: diagrama de dispersión y estudio de capacidad de proceso.

En la Figura 2, se muestra el diagrama de dispersión Orificio 1 vs Orificio 2 donde se observa la no existencia de correlación al no presentarse patrón entre las dos variables, y se puede decir que las variables tienden a ser independientes. Existen muy pocos puntos que están dentro del área de especificación, es decir, donde ambos orificios cumplen con la especificación de calidad y son piezas aceptables.
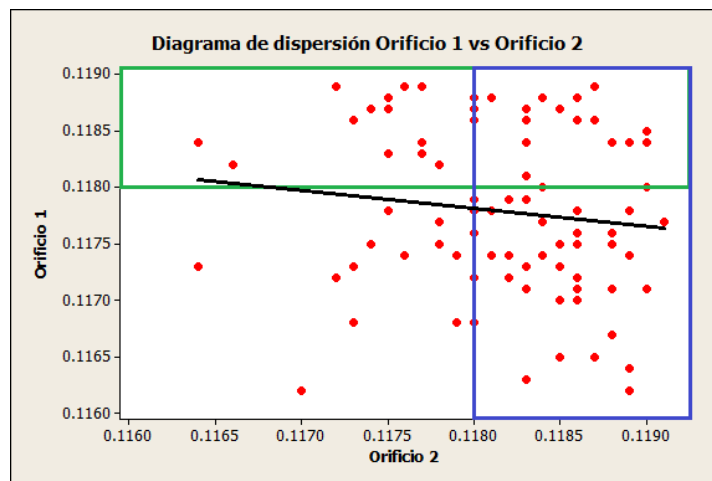

Figura 2. Diagrama de dispersión Orificio 1 vs Orificio 2

Además, se realizó un análisis de capacidad de proceso para ambos orificios y de acuerdo al gráfico obtenido se puede concluir lo siguiente: el comparativo que se realiza respecto a las especificaciones del producto nos indica que existe gran parte de las piezas que no cumplen con las especificaciones; el índice tanto de C_p como de C_{pk} refuerzan lo anterior al obtenerse valores de C_p= 0.91 y C_{pk}= -0.10 para el Orificio 1 de C_p= 1.00 y C_{pk}= 0.08 para el Orificio 2; las partes por millón totales de no conformidad es de 611,111 piezas para el Orificio 1 y 300,000 piezas para el Orificio 2, las cuales están por debajo de la especificación inferior.

Elección de los Factores, los Niveles y los Rangos

El criterio utilizado para establecer los factores tipo constantes o variables se llevó a cabo por medio de una entrevista con el Ingeniero de Moldeo, especialista y responsable del

proceso de moldeo. En la Tabla 1 y 2 se muestran los factores que intervienen en el proceso y se clasifican en factores constantes o variables.

Tabla 1. Factores constantes que intervienen en el proceso de moldeo

Factores constantes	Valor Constante
Temperatura ambiental	18° C
Temperatura del barril	0° F
Tiempo de expulsión	1 seg
Presión del clamp	.90
Presión del colchón	1.00
Velocidad de cierre	2.85
Velocidad del tornillo	1
Pines utilizados	2 piezas
Fixtures	2 herramientas

Tabla 2. Factores variables que intervienen en el proceso de moldeo

Factores variables	Rango Inferior	Rango Superior
Temperatura por arriba del molde	345°F	375°F
Temperatura por abajo del molde	345°F	375°F
Tiempo de inyección	1 min	3 min
Tiempo de enfriamiento	1 min	3 min
Velocidad de inyección	1.5	2.5
Presión	1 bar	1.8 bar

Selección de la Variable de Respuesta
El diámetro de los orificios 1 y 2 de la pieza 1-1532082-9 deben estar dentro del rango de especificación de 0.118 a 0.122 mm que se exige de acuerdo al diseño y plano de la pieza.

4.2   Elección del Diseño de Experimento

Para analizar el problema que se tiene en el proceso de moldeo por inyección, se realiza un primer diseño de experimentos llamado "corrida de exploración". El diseño es un experimento 2^3 y contiene 5 corridas que se realizan en los puntos centrales de los parámetros.

4.3   Análisis Estadístico de los Datos

Después de realizar el primer diseño de experimentos o corrida de exploración, se midió el diámetro de los orificios 1 y 2 de cada pieza. Se utilizó el software Minitab para analizar los datos y realizar un Análisis de Varianza para cada orificio y se observa que no se muestra ningún parámetro significativo ya que $P \geq 0.05$, por lo que se realizó un análisis más detallado del proceso para detectar algún otro factor que no se estaba

tomando en cuenta. Se analizaron las herramientas que se utilizan y se revisó la máquina internamente y se detectó la utilización de un fixture llamado PIN, el cual es insertado dentro del molde y las piezas para después iniciar el proceso de moldeo por inyección.

Figura 3 Fixtures "PIN" utilizado durante el proceso de moldeo de la pieza

Para analizar el fixture "PIN" se midió la altura de los orificios de la pieza 1-1532082-9 con el vernier Mitutoyo absolute. El resultado obtenido nos muestra que la altura de los orificios es de 0.1940 mm, después se mide el grosor de los pines que se utilizan durante el proceso, para de esta manera analizar si puede ser un factor crítico sobre las dimensiones del diámetro de los orificios del silicón. Se observó que a una altura de 0.1940 mm se detecta un grosor en el PIN de 0.1172 mm, por lo cual se considera como una causa potencial que puede afectar al proceso productivo en relación a las especificaciones de calidad del orificio 1 y 2.

Al considerar los PIN como un factor potencial, se propuso realizar unos pines que a una altura de 0.1940 mm muestre un grosor de 0.120 mm y posteriormente realizar un diseño de experimentos para verificar su implementación. El diseño que se utilizó para determinar si el grosor del PIN es un factor crítico en el diámetro de los orificios y se realizó un experimento con un solo factor donde se hace variar el "PIN" utilizado en el proceso. Se clasificaron los pines en Pin A y Pin B, donde el Pin A es el par que muestra un grosor de 0.1172 mm y e Pin B es el par que muestra un grosor de 0.120 mm

De acuerdo a los datos recolectados de cada muestra, se obtiene los resultados mostrados en la Tabla 3.

Tabla 3. Análisis estadístico de los resultados del diseño de experimentos con un solo factor para el Orificio 1 y Orificio 2

Orificio 1	Pin A	Pin B
Media	0.117	0.119
StDev	0.001	0.001
Media Mínima	0.116	0.119

Orificio 2	Pin A	Pin B
Media	0.119	0.121
StDev	0.000	0.000
Media Mínima	0.118	0.120

Media Máximo	0.118	0.120
Valor Mínimo	0.115	0.118
Valor Máximo	0.120	0.121

Media Máximo	0.119	0.121
Valor Mínimo	0.118	0.120
Valor Máximo	0.120	0.122

Prueba de Diferencia de Medias para Orificio 1 y Orificio 2

1. Se realiza una prueba de diferencia de medias para el Orificio y Orificio 2 antes y después de la modificación del Pin.

Ho: La   medias   antes   y   después   son   iguales
H1: La medias después del cambio es superior a la media antes del cambio

Tabla 4. Prueba de diferencia de medias para Orificio 1

```
            N      Mean    StDev  SE Mean
Orificio_1 10  0.119440  0.000750  0.00024
Orificio   10   0.11725   0.00133  0.00042

Difference = mu (Orificio_1) - mu (Orificio)
Estimate for difference:  0.002190
95% CI for difference:  (0.001175, 0.003205)
T-Test of difference = 0 (vs not =): T-Value = 4.53   P-Value =
0.000  DF = 18
Both use Pooled StDev = 0.0011
```

Tabla 5. Prueba de diferencia de medias para Orificio 1

```
            N      Mean    StDev  SE Mean
Orificio_2 10  0.120760  0.000488  0.00015
Orificio   10  0.118610  0.000412  0.00013

Difference = mu (Orificio_2) - mu (Orificio)
Estimate for difference:  0.002150
95% CI for difference:  (0.001726, 0.002574)
T-Test of difference = 0 (vs not =): T-Value = 10.64   P-Value =
0.000  DF = 18
Both use Pooled StDev = 0.0005
```

Como se muestra en la Tabla 4 y 5 se rechaza la hipótesis nula, es decir, la media después del cambio se incrementó.

4.4   Implementación de Resultados

Para confirmar que los diámetros del Orificio 1 y 2 cumplen con las especificaciones de calidad, se tomaron 30 mediciones de un día de producción utilizando el par de Pin B de 0.120 mm de grosor. Los datos obtenidos se concentran estadísticamente en la Tabla 6.

Tabla 6. Análisis estadístico de los datos obtenidos para Orificio 1 y Orificio 2 con Pin B

Orificio 1			Orificio 2	
	Pin B			**Pin B**
Media	0.119		**Media**	0.120
StDev	0.0003		**StDev**	0.0003
Media Mínima	0.119		**Media Mínima**	0.120
Media Máximo	0.120		**Media Máximo**	0.121
Valor Mínimo	0.119		**Valor Mínimo**	0.12
Valor Máximo	0.120		**Valor Máximo**	0.121

Como se pude observar, la media de ambos orificios se encuentran dentro de especificación, así como el valor mínimo y máximo de cada uno, por lo tanto se concluye que el proceso se encuentra dentro de especificación.

5   Resultados

Al determinar que el Pin era un factor crítico que afecta en la variable de respuesta, se realizó una modificación en el diseño, es decir, a una altura de 0.194 mm el grosor tiene que ser 0.120 mm. Se puede resaltar que en el estado inicial del proceso las partes por millón totales de no conformidad eran de 611,111 piezas para el orificio 1 y 300,000 piezas para el orificio 2, las cuales se localizaban por debajo de la especificación inferior. Después de rediseñar el fixture, las partes por millón totales de no conformidad disminuyeron en su totalidad a 0 piezas fuera de especificación para ambos orificios. Además, el índice tanto de C_p como de C_{pk} mejoró en comparación de los resultados obtenidos anteriormente para cada orificio, es decir, la capacidad potencial incrementó de un $C_p = 0.91$ y $C_{pk} = -0.10$ a $C_p = 1.66$ y $C_{pk} = 1.17$ para el orificio 1 y de un $C_p = 0.1$ y $C_{pk} = 0.08$ a $C_p = 1.83$ y $C_{pk} = 1.51$ para el orificio 2.

6   Conclusiones

El diseño de experimentos es altamente efectivo para aquellos procesos, que su rendimiento se ve afectado por varios factores. Con esta técnica se puede conseguir entre otras, mejorar el rendimiento de un proceso, reducir su variabilidad o los costos de producción.

Al realizar un diseño de experimentos en el proceso de moldeo por inyección de la pieza 1-1532082-9 se concluye que el factor principal que interviene en la calidad del producto es un par de fixture llamados "Pin". Después de rediseñar el fixture se obtuvieron mejores resultados en el proceso de produccion teniendo 0 piezas fuera de especificación para ambos orificios. Al mejorar el proceso se obtiene un impacto positivo en la satisfacción del cliente, cumpliendo sus expectativas de manera mas efectiva y

eficiente, al embarcar un producto que cumple con las características de calidad, además, se eliminan los tiempos de retrabajo, y se minimizan los costos por factor de scrap o desperdicio.

Referencias

1. Camisión, C., and Pérez, D. J.(2010). The future of the quality/excellence function: A vision from the Spanish firm. Routledge, 649-672
2. Shuaib, N., Nasir, S., Fathullah, M., Shayfull, Z., & Abdul Manan, M. (2012). The Influence of Different Mold Temperature on Warpage in a Thin Shallow Injection Molding Process. International Review of Mechanical Engineering, 11-16
3. Kurt, K., Kaynak, A., & Girit, O. (2009) Experimental investigation of plastic injection molding: assessment of the effects of cavity pressure and mold temperature on the quality of the final products. Master Des, 3217-3224
4. Curic, D., Veljkovic, Z., & Duhovnik, J. (2012). Comparison of methodologies for identifícation of process parameters affecting geometric deviations in plastic injection molding of housing using Taguchi method. Mechanika, 671-676.
5. Saurav, D., Bandyopadhyay, A., & Kumar, P. (2008). Greybased Taguchi method for optimization of bead geometry in submerged arc bead-on iplate. International Journal of Advance Manufacturing Technology, 1136-114.
6. Ko-Ta, C., & Fu-Ping, C. (2006). Analysis of shrinkage and warpage in an injection-molded part with a thin shell feature using the response surface methodology. Int J Adv Manuf Technol, 468–479
7. Jeang, A. (1999). Robust Tolerance Desing by Response Surface Methodology. Advance Manufacturing Technology, 399-403
8. Antony, J., & Antony, J. Teaching the Taguchi method to industrial engineers. Work Study, 141-149 (2001).
9. Ming-Tsan, C., Yung-Kuang, Y., & Yun-Hsiang, H. Modeling and Optimization of Injection Molding Process Parameters for Thin-Shell Plastic Parts. Polymer-Plastics Technology & Engineering, 745-753 (2009).
10. Montgomery, D. (2005). Diseño y Análisis de experimentos. Limusa S.A de C.V.

ANÁLISIS CRÍTICO DEL DISEÑO FACTORIAL 2^k SOBRE CASOS APLICADOS

Critical analysis of 2k factorial design based on applied cases

RESUMEN

En éste artículo se exponen las definiciones de diseño experimental y específicamente del diseño factorial 2^k, su importancia, características, ventajas y desventajas; además de presentar tres casos aplicados de dicho diseño con su correspondiente análisis, explicando por qué en cada uno de estos casos fue apropiado aplicar el método factorial 2^k. Se muestran también los parámetros bajo los cuales es adecuado utilizar este método y se concluye que es uno de los métodos más convenientes de aplicar cuando se requiere analizar los efectos conjuntos de varios factores y las interacciones entre ellos.

PALABRAS CLAVES: Diseño experimental, diseño factorial 2^k, interacciones, tratamientos, variables.

ABSTRACT
This paper presents the definition of experimental design, specifically the one for 2k factorial design, together with its importance, characteristics, advantages and disadvantages. Three applied cases of 2k factorial design are analyzed and it is explained why for each one of them was appropriated to apply this method. It is shown what parameters must be considered to choose this experimental design and we concluded that it is one of the more convenient to apply when we need to analyze the combined effects of various factors and the interactions between them.

KEYWORDS: *Experimental design, 2k factorial design, interactions, treatment, variables*

PEDRO DANIEL MEDINA VARELA
Profesor Asistente
Facultad de Ingeniería Industrial
Universidad Tecnológica de Pereira
pemedin@utp.edu.co

ANGELA MARIA LOPEZ REYES
Docente Universidad Antonio Nariño – Sede Roldanillo
angela.lopez@uan.edu.co

1. INTRODUCCIÓN

Los diseños factoriales son ampliamente utilizados en experimentos en los que intervienen varios factores para estudiar el efecto conjunto de éstos sobre una variable de interés. Existen varios casos especiales del diseño factorial general que resultan importantes porque se usan ampliamente en el trabajo de investigación, además de constituir la base para otros diseños de gran valor práctico.

Uno de los más importantes de estos casos especiales ocurre cuando se tienen k factores, cada uno con dos niveles. Estos niveles pueden ser cuantitativos como sería el caso de dos valores de temperatura, presión o tiempo pero también pueden ser cualitativos como sería el caso de dos máquinas, dos operadores, los niveles "superior" e "inferior" de un factor, o quizás, la ausencia o presencia de un factor. Una réplica completa de tal diseño requiere que se recopilen 2 x 2 x .... x 2 = 2^k observaciones y se conoce como diseño general 2^k.

2. DISEÑO ESTADÍSTICO DE EXPERIMENTOS

Como se menciona en [1], se podría definir el diseño estadístico de experimentos también denominado diseño experimental, como una metodología basada en herramientas matemáticas y estadísticas cuyo objetivo es ayudar al experimentador a:

- Seleccionar la estrategia experimental óptima que permita obtener la información buscada con el mínimo costo.
- Evaluar los resultados experimentales obtenidos, garantizando la máxima fiabilidad en las conclusiones que se obtengan.

El diseño experimental es aplicado ampliamente al estudio de los procesos de producción. Un proceso puede considerarse como una "caja negra" a la cual ingresan diversas variables que interactúan para producir un resultado. Las variables que ingresan al proceso se denominan variables de entrada, y el resultado, variable de salida. El nivel de la variable de salida depende de los niveles que adopten las variables de entrada, y los gerentes y técnicos se benefician al saber qué combinación de variables de entrada produce el mejor nivel en la variable de salida. La búsqueda de

combinaciones óptimas de las variables de entrada da lugar al diseño experimental, que es una prueba (o un conjunto de pruebas) durante la cual se realizan cambios sistemáticos y controlados a las variables de entrada para medir el efecto sobre la variable de salida. [2]

2.1 Importancia del diseño experimental

El principal propósito de la experimentación es obtener información de calidad, por lo que el papel que juega la experimentación en todos los campos de la investigación y el desarrollo es primordial. Dicha información es la que permite el desarrollo de nuevos productos y procesos, la mejora en cuanto a la comprensión de un sistema (ya sea un proceso industrial, un procedimiento analítico, entre otros) para así poder tomar las decisiones correctas acerca de cómo optimizarlo y mejorar su calidad, comprobar hipótesis científicas, etc.
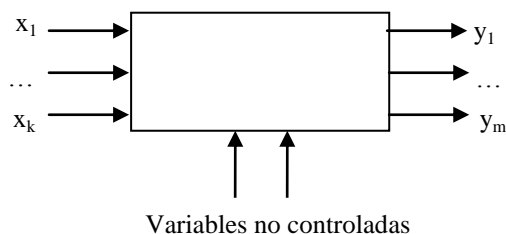

Figura 1. Representación de un sistema en estudio en diseño estadístico de experimentos: factores (x), respuestas (y). [1]

3. DISEÑO FACTORIAL [3]

Para optimizar procesos de fabricación, condiciones de reacción y métodos de análisis entre otros, es necesario conocer qué variables influyen significativamente en el sistema y como lo afectan. A menudo esta información no está disponible y se genera experimentando. Primero se recogen en una lista todas las variables que podrían influir en la respuesta[1]. A continuación, se realiza una serie de experimentos en los cuales se fijan las variables que son de interés modificar, se anota el valor de las que no se pueden controlar, y se varían las restantes. Finalmente, se obtiene la información comparando la variación de la respuesta entre experimentos.

El método tradicional de variar un factor cada vez no suele ser la mejor opción. Puede implicar más experimentos de los necesarios y, a pesar de ello, proporcionar sólo información parcial [10]. Por ejemplo, no mostrará si existe interacción entre factores. Las interacciones suelen ser muy corrientes y a veces son los efectos más importantes, por lo que conocerlas es imprescindible para comprender el comportamiento de muchos sistemas.

El diseño estadístico de experimentos contempla una amplia variedad de estrategias experimentales que son adecuadas para generar la información que se busca, el **diseño factorial completo 2^k** es una de ellas. Este describe los experimentos más adecuados para conocer simultáneamente qué efecto tienen k factores sobre una respuesta y descubrir si interaccionan entre ellos.

Estos experimentos están planeados de forma que se varían simultáneamente varios factores pero se evita que se cambien siempre en la misma dirección. Al no haber factores correlacionados se evitan experimentos redundantes. Además, los experimentos se complementan de tal modo que la información buscada se obtiene combinando las respuestas de todos ellos. Esto permite obtener la información con el mínimo número de experimentos (y por tanto, con el menor coste) y con la menor incertidumbre posible (porque los errores aleatorios de las respuestas se promedian).

Un experimento factorial permite investigar todas las combinaciones posibles de los niveles de los factores o condiciones en cada prueba completa. Los tratamientos constan de todas las combinaciones que puedan formarse de los distintos factores, entendiéndose como tratamientos la denominación de los diferentes procesos cuyos efectos van a ser medidos o comparados. El objetivo es investigar los resultados experimentales en casos donde interesa estudiar el efecto de diversas condiciones de experimentación y sus interacciones. [4]

4. CASOS APLICADOS DEL DISEÑO FACTORIAL 2^k

El diseño de experimentos encuentra numerosas aplicaciones en el campo de las reacciones químicas e industrias farmacéuticas. En este artículo se hará referencia a tres casos particulares donde el diseño experimental factorial 2^k fue utilizado básicamente para conocer las interacciones entre los diferentes factores que intervinieron en cada experiencia y determinar cuál es la forma óptima de manejar dichos factores.

4.1 Primer caso

"Procesos de oxidación avanzada (AOPs) aplicados en la degradación de oxitetraciclina: un antibiótico persistente en efluentes industriales acuícolas" [5]

Un resumen de la investigación encabezada por el ingeniero M. Ollino [5] es: "En este trabajo se estudió la degradación oxidativa del antibiótico oxitetraciclina (**OT**) de uso común en la industria salmonera chilena, con el propósito de evaluar la potencialidad de los procesos de oxidación avanzada, AOPs, en la destrucción de agentes recalcitrantes presentes en aguas de la industria salmonífera. La característica que define a los

---

[1]Respuesta es la variable de interés que se mide como consecuencia de la experimentación.

AOPs es la generación de especies radicales que son poderosos oxidantes en solución. La oxidación de **OT** se llevó a cabo tratándolo con ozono, una combinación ozono-UV y una combinación ozono-peróxido, en reactores fotoquímicos previamente diseñados en nuestros laboratorios.

El objetivo principal es: usando un diseño factorial, determinar las mejores condiciones experimentales para la eliminación vía oxidación del antibiótico oxitetraciclina (**OT**) usando los siguientes sistemas: **(i)** combinación de ozono y radiación UV y **(ii)** combinación de ozono con H_2O_2 a distintos valores de pH. La función objetivo para cada corrida fue el porcentaje remanente del antibiótico tras 50 minutos de tratamiento (t_{50}), es decir, una forma alternativa de evaluar la eficiencia energética de cada reacción, medido como el consumo de energía eléctrica en función del tiempo, necesario para producir el ozono consumido durante la eliminación del polutante."

Para éste caso, el diseño factorial 2^k fue muy conveniente, debido a que éste permitió:

➢ Determinar las mejores condiciones experimentales para eliminar la oxitetraciclina (**OT**), oxidándola a distintos valores de pH.

➢ Disminuir el número de experimentos que normalmente deben realizarse, permitiendo evaluar el peso relativo de cada variable analizada (pH, concentración, sistema oxidante utilizado, presencia de aceptores de electrones, longitud de onda de irradiación, etc.) y establecer los posibles efectos sinérgicos o antagónicos entre las variables.

➢ Analizar la degradación oxidativa OT con un punto central, para comparar simultáneamente el efecto de las variables PH inicial y sistema oxidante utilizado consideradas sobre éste fenómeno.

4.2 Segundo caso

"Diseño factorial 2^k aplicado a la determinación del valor de conservación de inmuebles de interés cultural". [6]

"Este artículo presenta un caso particular del diseño de experimentos usado para determinar el valor de conservación de los inmuebles de interés cultural en el municipio de Apia (Risaralda - Colombia), permitiendo comprobar que las variables con mayor incidencia en la determinación del valor de conservación son: valor de construcción, área del terreno, valor del terreno y una variable generadora. El uso del diseño de experimento factorial 2^k permitió evidenciar que en la determinación

del valor de los inmuebles de interés cultural, es posible reducir el número de variables si estas aportan mayor significancia para el cálculo del valor de conservación de los inmuebles".

El diseño factorial 2^k fue utilizado en este estudio por dos razones fundamentales:

➢ Se requería determinar cuál de las variables consideradas en el avalúo de los inmuebles caracterizaban mejor el valor de conservación de los inmuebles de interés cultural ubicados en el municipio de Apia, Risaralda.

➢ Era inminente determinar el valor de conservación de los inmuebles tomando un número reducido de variables, es decir las que se consideran más representativas.

4.3 Tercer caso

"Formulación de una mezcla que disminuya la tensión superficial del agua para uso en la agricultura comercial". [7]

Éste título corresponde a un trabajo de grado realizado por la tecnóloga en química industrial Aura Arias en 2005, donde plantea que: "la necesitad de obtener un coagulante para el uso de la agricultura comercial que cumpla con la función de disminuir la tensión superficial del agua, humectar las plantas, que sea fácilmente biodegradable y compatible con la estructura de las hojas de las plantas a las cuales van a ser aplicadas, utilizando como instrumento de medida un estalagmómetro, el cual relaciona la densidad, el peso y el volumen de la gota que desciende por un capilar en función de la tensión superficial, se consiguió por medio de un diseño factorial".

La utilización del diseño factorial en esta investigación contribuyó a la evaluación de varias moléculas surfactantes y humectantes comerciales, permitiendo determinar los rangos de concentración dentro de los cuales se obtiene la máxima caída en la tensión superficial con la mínima aplicación de agente tensoactivo. Además, el diseño factorial permitió también conocer las interacciones entre los diferentes tipos de moléculas y determinó la viabilidad de la aplicación sola o conjunta de varias de ellas.

El diseño factorial implementado hizo posible predecir matemáticamente la tensión superficial para las mezclas evaluadas, así mismo el cálculo del valor estimado y el valor observado predice con mucha exactitud la tensión superficial; lo anterior permitió a su vez evaluar dentro de los rangos de concentración de moléculas escogidos, mezclas particulares para su uso en aplicaciones especiales ó más específicas.

5. ANÁLISIS DE CASOS

Comúnmente para optimizar un protocolo que posee distintas variables se emplea un método lógico, el cual consiste en mantener constantes k-1 variables y cambiar paulatinamente la variable restante hasta encontrar un punto donde su composición sea la más óptima. Éste método no es recomendable para ninguno de los tres casos expuestos anteriormente, puesto que no tiene en cuenta las interacciones conjuntas de las variables, las cuales se presentan muy a menudo en experimentos que poseen pocas variables y experiencias donde unas variables dependen de otras.

Para tener en cuenta las interacciones de las diferentes variables de los tres casos ya mencionados, se hizo uso de un modelo matemático estadístico conocido como diseño factorial 2^k.

Los diseños factoriales 2^k, ó también conocidos como planes de experiencia son ampliamente utilizados en experimentos donde se requiere encontrar valores óptimos de un grupo de variables, las cuales están determinadas por un rango de valores, éstos se conocen como niveles de la variable. Estos niveles pueden ser cuantitativos (primer y tercer caso), cuando se trata de niveles numéricos ó cualitativos (segundo caso) cuando se trata la ausencia o presencia de un factor.

Para los tres casos anteriores, el diseño factorial 2^k, como su nombre lo indica, manejó dos niveles, presentando entonces un nivel superior y un nivel inferior, los cuales pueden ser denotados (como lo recomienda Montgomery [8]) como (+) y (-) respectivamente. Esto con el fin de construir la matriz de ensayos que indica los niveles en los que se deben realizar las corridas. A modo de ejemplo, para el tercer caso dicha matriz se construyó con tres variables, como se muestra en la siguiente tabla:

Nombre de la variable	Convención	Nivel inferior	Nivel superior
Molécula 1	A	-	+
Molécula 2	B	-	+
Molécula 3	C	-	+

Tabla 1: Matriz de niveles máximos (+) y mínimos (-) de las variables del diseño experimental del tercer caso. [7]

Ésta metodología tiene las ventajas de: no desperdiciar materia prima (reactivos, aditivos, entre otros), requiere poco tiempo de análisis, tiene en cuenta las interacciones bifactoriales, trifactoriales, etc, también posee la gran ventaja de que una vez realizado el plan de experiencias, se obtiene una ecuación que describe el comportamiento del sistema dentro de los rangos predichos.

En los tres casos observados, se aplican dos de las características fundamentales del diseño experimental como lo son:

- ➤ Manipulación: es la intervención deliberada del investigador para provocar cambios en la variable dependiente.

- ➤ Aleatorización: es la manera como se distribuyen los sujetos en los diferentes grupos que forman parte del estudio.

La utilización del diseño experimental factorial 2^k fue muy favorable y sobre todo adecuada para llevar a cabo las investigaciones ya mencionadas, puesto que en éstas era de vital importancia estudiar los efectos que producen una o más variables, sus consecuencias y la relación causa-efecto que se puede originar entre ellas, parámetros bajo los cuales el diseño factorial cobra su importancia, dado que son precisamente éstos los argumentos por los cuales se determina la utilización o no de dicho diseño experimental.

Otro parámetro fundamental dentro de cualquiera que sea la experimentación a realizar, es conocer el error y reducirlo todo lo posible, por ello los estudios de investigación deben ser y reunir la característica de la aleatorización, con el objetivo de distribuir de manera uniforme cualquier efecto de un factor incontrolable.

Emplear el diseño factorial 2^k en los tres casos señalados con anterioridad trajo consigo muchas ventajas como:

- ➤ Eliminó el efecto de las variables perturbadoras o extrañas mediante el efecto de aleatorización.

- ➤ Controló y manipuló los factores controlables, para poder clarificar la dirección y naturaleza de las causas estudiadas.

- ➤ Permitió la flexibilidad, eficiencia, simetría y manipulación estadística.

5.1 Calidad del diseño experimental

Con respecto a la calidad del diseño factorial 2^k, se puede decir que éste cuenta con:

- ➤ **Validez interna:** Es el grado en que los cambios observados se pueden atribuir a la manipulación experimental. Estudia hasta que punto una causa puede ser atribuida a un efecto. Sin embargo, hay que resaltar que cuantas más variables entran en un diseño se va restando validez interna. De esta forma, se puede afirmar que para los casos analizados, el diseño factorial 2^k fue utilizado apropiadamente ya que las

---

[1]Respuesta es la variable de interés que se mide como consecuencia de la experimentación.

Scientia et Technica Año XVII, No 47, Abril de 2011. Universidad Tecnológica de Pereira.

105

variables que entraban en juego no eran muy numerosas y por el contrario, se trató mediante la aplicación del mismo minimizar dichas variables.

➢ **Validez externa:** es el grado en que los resultados de un estudio pueden ser generalizados a muestras o condiciones espacio – temporales diferentes. Por ejemplo: "A" causa "B", pero seguiría causando "B" con otros: sujetos, contexto (validez ecológica), momentos.

➢ **Validez ecológica:** Es la que se puede aplicar en distintos contextos. En los tres casos quedó evidenciado que el diseño factorial 2^k puede ser utilizado para una diversidad de aplicaciones sin que se generen alteraciones en su metodología ó en la confiabilidad y eficiencia de los resultados obtenidos: cada caso es completamente diferente a los demás y en cada uno de ellos se logró alcanzar el objetivo deseado.

Aunque ya se ha dicho desde el inicio de este artículo, es importante hacer énfasis en la característica principal del diseño factorial 2^k y que se vio ampliamente demostrada en los tres casos vistos; dicha característica es que el diseño factorial permite observar el comportamiento de una variable bajo todas las condiciones. Permite valorar el efecto de interacción (el efecto combinado de ambas variables), es decir, permite saber el efecto principal de A, el de B y el efecto combinado de ambos).

5.2 Ventajas de la experimentación factorial cuando los factores son independientes [4]

Las ventajas de la experimentación factorial descrita, dependen naturalmente de la finalidad del experimento. Supóngase por ahora que, que el propósito es investigar los efectos de cada factor, sobre algún intervalo pre asignado que está cubierto por los niveles de ese factor usados en el experimento. En otras palabras, el objeto es obtener un cuadro amplio de los efectos de los factores, más bien que encontrar, por ejemplo, la combinación de los niveles de los factores que dan una respuesta máxima (que era el objetivo principal en el primer, segundo y tercer caso). Un procedimiento para esto es conducir experimentos separados, cada uno de los cuales considerará un solo factor (pero ya se ha visto que éste método no es nada eficiente). Otro procedimiento es incluir todos los factores simultáneamente por medio de un experimento factorial.

Si todos los factores son independientes en sus efectos, el método factorial significará un ahorro considerable de tiempo y material dedicado a los experimentos. El ahorro se deriva de dos hechos: primero, cuando los factores son independientes todos los efectos simples de un factor son iguales a su efecto principal, de tal manera que los

efectos principales son las únicas cantidades necesarias para describir completamente las consecuencias de las variaciones en el factor. Segundo, en un experimento factorial cada efecto se estima con la misma precisión que si todo el experimento se hubiese dedicado a ese solo factor. Por lo tanto, si hubiese k factores únicos, todos a dos niveles y todos independientes, el método del factor único necesitaría k veces el material experimental que requeriría un arreglo factorial de la misma precisión. La ganancia obtenida de los arreglos factoriales, en este caso es bastante substancial.

Las consideraciones de tipo práctico pueden disminuir esta ganancia. El experimentador frecuentemente carece de los recursos para conducir un experimento grande con varios tratamientos y debe trabajar únicamente con uno o dos factores a la vez. Además ya se ha hecho notar con anterioridad que, conforme aumenta el número de combinaciones de tratamientos en un experimento, el error estándar por unidad también aumenta. Por lo tanto éste error es más probable que sea más alto para un experimento factorial grande, que para un experimento similar con uno o dos factores como el 2^k.

5.3 Ventajas de la experimentación factorial cuando los factores no son independientes. [4]

Se supone que el propósito es, como en la sección anterior, investigar cada factor en el intervalo representado por sus niveles. Cuando los factores no son independientes, los efectos simples de un factor varían de acuerdo con la combinación particular de los otros factores, con los cuales éstos se producen. En este caso, es probable que el método del factor único provea solamente un número de fragmentos inconexos de información, que no puedan ser ligados fácilmente. Para conducir un experimento con un factor único A, debe tomarse alguna decisión acerca de los niveles de los otros factores, por ejemplo B, C y D, que se van a usar en el experimento. El experimento revela, entonces, los efectos de A para esta combinación particular de B, C y D; pero no da ninguna información para predecir los efectos de A con cualquier otra combinación de B, C y D. Por otro lado, con el método factorial, los efectos de A se examinan con todas las combinaciones de B, C y D, que se incluyen en el experimento. Así, se acumula una gran cantidad de información sobre los efectos de los factores y sobre sus relaciones entre sí.

6. CONCLUSIONES Y RECOMENDACIONES

Los diseños factoriales completos son la estrategia experimental óptima para estudiar simultáneamente el efecto de varios factores sobre la respuesta y sus interacciones. Por su potencia y sencillez, su campo de aplicación es muy amplio:

➢ Identificar que variables influyen en una reacción, para luego poder optimizarlas hasta alcanzar el rendimiento deseado, o para disminuir el tiempo de reacción.

➢ Decidir que se debe ajustar en el nuevo proceso de fabricación para que no se produzcan tantos productos fuera de especificaciones.

➢ Estudiar en qué condiciones el proceso es más robusto a pequeñas variaciones de temperatura, humedad…

Cuando los factores que se va a investigar son numerosos, la principal desventaja del experimento factorial 2^k estriba en su tamaño y complejidad. Sin embargo, la magnitud de la tarea puede reducirse teniendo en cuenta únicamente una sola repetición para cada tratamiento. Hasta es posible obtener casi toda la información deseada probando solamente una fracción (por ejemplo, la mitad o un cuarto) del número total de combinaciones de tratamientos, aun cuando esto se haga con ciertos riesgos.

Las dificultades que surgen de los experimentos grandes no deben ser consideradas como una crítica del método factorial 2^k, puesto que la base de la dificultad simplemente es que se ejecuta un programa de investigación más grande. Para resumir, a continuación se dan algunos casos donde la experimentación factorial puede ser adecuada:

➢ En trabajos de exploración, donde el objeto es determinar rápidamente los efectos de cada uno de cierto número de factores dentro de un intervalo específico.

➢ En investigaciones de las interacciones entre los efectos de varios factores. Por su naturaleza las interacciones no se pueden estudiar sin probar algunas de las combinaciones que se forman de los diferentes factores. Frecuentemente la información se obtiene mejor probando todas las combinaciones.

➢ En experimentos diseñados para poder llegar a recomendaciones que deben aplicarse a una gran variedad de condiciones. Se pueden introducir factores auxiliares en un experimento para probar los factores principales bajo una variedad de condiciones similares a las encontradas en la población a la cual se va a aplicar dichas recomendaciones.

7. REFERENCIAS BIBLIOGRÁFICAS.

[1] J. Ferré y X. Rius, "Introducción al diseño estadístico de experimentos," *Técnicas de Laboratorio*, vol. 24, No. 274, pp. 648-653, Septiembre de 2002.

[2] E Yacuzzi, F. Martín, H. M. Quiñones y M. J. Popovsky, "El diseño experimental y los métodos de Taguchi: conceptos y aplicaciones en la industria farmacéutica," *Pharmaceutical Management*, pp 1-28, Febrero de 2004

[3] J. Ferré y X. Rius, "Diseño factorial completo 2^k," *Técnicas de Laboratorio*, vol. 25, No. 287, pp. 999-1004, Octubre de 2003.

[4] W. G. Cochran y G. M. Cox, Diseños experimentales. México: Editorial Trillas, 1990, cap 5, pp. 177-182.

[5] M. Ollino, C. Lizama, C Caneo y P Pérez. "Procesos de oxidación avanzada (AOPs) aplicados en la degradación de oxitetraciclina: un antibiótico persistente en efluentes industriales acuícolas". *Scientia et Technica,* Año XIV, No 39, Universidad Técnica Federico Santa María. Chile, Julio de 2008.

[6] C. O. Lozada Riascos y J. A. Urrutia Mosquera. Diseño factorial 2^k aplicado a la determinación del valor de conservación de inmuebles de interés cultural. Universidad Tecnológica de Pereira. Colombia, Septiembre de 2008.

[7] A. M. Arias Gonzales. Formulación de una mezcla que disminuya la tensión superficial del agua para uso en la agricultura comercial. Universidad Tecnológica de Pereira. Colombia, 2005.

[8] D. Montgomery. Diseño y análisis de experimentos. Editorial Iberoamericana, Estados Unidos. 1991, pp. 589.

[9] A. Muñoz Escalona Pavón. "Control de Calidad". Máster de Organización Industrial y Gestión de Empresas. (2008, Noviembre 24).

[10] CRUZ, Eduardo; RESTREPO, Correa y MEDINA, Pedro. Aplicación de un Modelo Factorial de Experimentación en un Ingenio Azucarero del Valle del Cauca. Revista Scientia et Technica. Año XII, No 32, pp 313 – 318. Diciembre de 2006.

---

[1]Respuesta es la variable de interés que se mide como consecuencia de la experimentación.

Review article

The application of analysis of variance (ANOVA) to different experimental designs in optometry

R. A. Armstrong, F. Eperjesi and B. Gilmartin

Vision Sciences, Aston University, Birmingham B4 7ET, UK

Abstract

Analysis of variance (ANOVA) is the most efficient method available for the analysis of experimental data. Analysis of variance is a method of considerable complexity and subtlety, with many different variations, each of which applies in a particular experimental context. Hence, it is possible to apply the wrong type of ANOVA to data and, therefore, to draw an erroneous conclusion from an experiment. This article reviews the types of ANOVA most likely to arise in clinical experiments in optometry including the one-way ANOVA ('fixed' and 'random effect' models), two-way ANOVA in randomised blocks, three-way ANOVA, and factorial experimental designs (including the varieties known as 'split-plot' and 'repeated measures'). For each ANOVA, the appropriate experimental design is described, a statistical model is formulated, and the advantages and limitations of each type of design discussed. In addition, the problems of non-conformity to the statistical model and determination of the number of replications are considered.

Keywords: Analysis of variance (ANOVA), experimental design, factorial experimental design, random effect factor, randomised blocks, repeated measures design, split-plot design.

Introduction

Analysis of variance (ANOVA) is the most efficient parametric method available for the analysis of data from experiments. It was devised originally to test the differences between several different groups of treatments thus circumventing the problem of making multiple comparisons between the group means using *t*-tests (Snedecor and Cochran, 1980). ANOVA is a method of great complexity and subtlety with many different variations, each of which applies in a particular experimental context. Hence, it is possible to apply the wrong type of ANOVA in a particular experimental situation and, as a consequence, draw the wrong conclusions from the data.

A previous article (Armstrong *et al.*, 2000) described the origin of ANOVA, the logic which underlies the method and the assumptions necessary to apply it to

data. In addition, the simplest form of the analysis, the one-way ANOVA in a randomised design, was described and illustrated using a clinical experiment drawn from optometric research. The various methods available for making specific comparisons between pairs of group means (also known as '*posthoc*' tests) were also considered.

The objective of the present article is to extend these techniques to different experimental designs. For the purpose of this article, these types of ANOVA will be considered under the following headings likely to cover many situations encountered in optometric research: (1) one-way ANOVA, 'random effects' model (2) two-way ANOVA in randomised blocks (3) three-way ANOVA (4) factorial ANOVA (5) factorial ANOVA, split-plot design, and (6) factorial ANOVA, repeated measures design. In each case, the type of experimental design is described, a statistical model is given and the advantages and limitations of the appropriate ANOVA discussed. In addition, the problems of non-conformity to the statistical model and determination of the number of replications are considered. The number of types of ANOVA described are illustrated with simple data sets drawn from clinical experiments in optometry. These data sets are used only to illustrate the methodology

and the results quoted may not be typical of those that would be observed in more extensive experiments.

Statistical models

In a previous article (Armstrong *et al.*, 2000), we described a commonly used notation to describe the basic model of an ANOVA. The subscript '*i*' is used to denote the group or class (i.e. the treatment group), '*i*' taking the values '1 to a', whereas the subscript '*j*' designates the members of the class, '*j*' taking the values '1 to n' (hence, '*a*' groups and '*n*' replicates per group). Within class '*i*', the observations x_{ij} are assumed to be normally distributed about a mean μ with variance s^2. This linear model can be written:

$$x_{ij} = \mu + a_i + e_{ij} \tag{1}$$

Hence, an observed value x_{ij} is the sum of three parts: (1) the overall mean of the observations (μ) (2) a treatment or class deviation '*a*', and (3) a random element '*e*' drawn from a normally distributed population. The random element reflects the combined effects of natural variation between replications and errors of measurement. All the more complex types of ANOVA can be derived from this simple model by the addition of one or more further terms to equation 1 (*Table 1*).

Types of ANOVA

One-way ANOVA, 'random effects' model

In our previous article (Armstrong *et al.*, 2000), we described a one-way ANOVA in a randomised design which compared the reading rates of three groups of subjects, viz., young normal subjects, elderly normal subjects and subjects with age-related macular degeneration. This ANOVA is described as a 'fixed effects' model in which the objective is to estimate the differences between the subject groups and these are regarded as 'fixed' or discrete effects. There is, however, an alternative model called the 'random effects' model in which the objective is to estimate the degree of variation of a particular measurement and in many circumstances to compare different sources of variation in space and time.

Example

An example of an experiment of this type is shown in *Table 2*. Five measurements of intraocular pressure (IOP) were made on a subject, 1 min apart on 3 days chosen randomly. The objective was to determine, for individual subjects, the degree of variation in IOP from minute-to-minute compared with the variation between measurements made on different days. In this case,

Table 1. Statistical models of the types of ANOVA described. For definition of terms see relevant text

Type of ANOVA	Y	Constant elements	Additional components	Interaction terms
One-way, fixed	x_{ij}	$\mu + e_{ij}$	a_i	
One-way, random	x_{ij}	$\mu + e_{ij}$	A_i	
Two-way, randomised blocks	x_{ij}	$\mu + e_{ij}$	$a_i + b_j$	
Three-way	x_{ijk}	$\mu + e_{ijk}$	$a_i + b_j + c_k + d_{ij}$	
Two-factor, factorial	x_{ijk}	$\mu + e_{ijk}$	$a_i + b_j$	$(ab)_{ij}$
Two-factor, factorial split-plot	x_{ijk}	$\mu + e_{ijk}$	$M_i + B_j + T_k + d_{ijk}$	$(MT)_{ik}$

In each case, the dependent variable (Y) can be considered to be the sum of the constant elements, the additional components and, where appropriate, the interaction terms in each row.

Table 2. The one-way ANOVA 'random effects model' (hierarchical or nested design) on five measurements of IOP made on a subject 1 min apart on 3 sample days (A)

Design Sample day	A1	A2	A3
Repeat Measurements	18	17	19
	19	18	18
	20	16	20
	19	17	20
	21	17	19

ANOVA table:

Variation	SS	d.f.	MS	F	ExpMS
Between days	17.73	2	8.866	10.64	$\sigma_m^2 + 5\sigma_D^2$
Between repeat measurements within days	10.0	12	0.833		σ_m^2

ExpMS = expected mean square. Components of variance: between days (σ_D^2) = 1.607, between repeat measurements (σ_m^2) = 0.833.

© 2002 The College of Optometrists

estimates of variability are the objective rather than the determination of fixed effects. Based on these estimates of variability, a suitable protocol for measuring IOP in a clinical context might be devised. For example, if IOP varied considerably from minute to minute but, on average, little between days, several measurements of the IOP on a single occasion might be an appropriate strategy. By contrast, if minute to minute variation was negligible but there was significant day to day variation, it might be better to measure IOP only once, but on a sample of days. This type of experiment may also be described as a 'hierarchical' or 'nested type' design, especially if each sample is composed of subsamples and these in turn are subsampled. In addition, variation may be spatial rather than temporal, e.g. visual function may be quantified at different locations on the retina. In this case, there may be variation between eyes, between different locations within the same retina, and between sequential measurements at the same location.

Model

Equation 1 describes a 'fixed effect' model in which the a_i are fixed quantities to be estimated. The corresponding 'random effects' model is similar (*Table 1*), but the symbol A_i, replaces a_i and represents the difference between the IOP of the ith patient and the mean. Hence, A_i is a random variable and the term e_{ij} refers to errors of measurement and to the fact that IOP is measured on a sample of occasions.

ANOVA

The ANOVA appropriate to this design provides an *F*-test of whether there is significant variation between days. When the null hypothesis is false, however, the mean square between classes (in this case between days) estimates the quantity, $\sigma_m^2 + n\sigma_D^2$ and it is possible to calculate the 'components of variance' σ_m and s_D from the ANOVA table (*Table 1*). These are estimates of the variance of the measurements made between days and between determinations on a single day. In the example quoted (*Table 2*), the *F*-test suggests significant variation

in IOP between days. In this context, however, the components of variance are more useful and indicate that the component between days is approximately twice as great as that on a single day.

It is important to identify whether a 'fixed' or 'random' effect model is appropriate in each case. This is particularly important in more complex factorial-type designs in which there may be a mixture of 'fixed' and 'random' effects (Snedecor and Cochran, 1980). One way of deciding whether a factor is 'fixed' or 'random' is to imagine the effect of changing one of the levels of the factor (Ridgman, 1975). If this makes it a different experiment, for example, by substituting a different group of subjects, then it is a 'fixed' effect factor. By contrast, if we considered it the same experiment, for example, by substituting a different sample day, as in the example described above, it would be a 'random' effect factor.

Two-way ANOVA *in randomised blocks*

In the one-way, 'fixed' effects ANOVA described previously (Armstrong *et al.*, 2000), each observation was classified in only one way, i.e. in which treatment or subject group the observation fell. Replicates were either allocated to treatment groups at random or subjects within a group were a random sample of a particular population. Such an experiment is often described as a 'randomised design'. More complex experimental designs are possible, however, in which an observation may be classified in two or more ways.

Example

An example of an experiment in which the observations are classified in two ways is shown in *Table 3*. This experiment studied the effect of four coloured filters on the reading rate of 12 patients grouped by age. Such an experiment is often called a 'randomised blocks design' because there is a restriction in how subjects are randomised to treatments: (1) subjects are first grouped into 'blocks' of similar age and (2) treatments are applied at random to the subjects within each block

Design	Red	Yellow	Green	Blue
Age 1	84.5	72.0	70.5	39.5
Age 2	79.3	68.2	62.6	47.2
Age 3	36.0	46.1	48.9	38.0
ANOVA table:				
Variation	**SS**	**d.f.**	**MS**	**F**
Treatments	1102.95	3	367.65	3.40
Blocks	1448.98	2	724.49	6.71*
Error	647.91	6	107.98	

*$p < 0.05$

Table 3. Two-way ANOVA in randomised blocks with four treatment groups (coloured filters) and with replicates also classified into three age groups (blocks). Data are the reading rate of the patient (number of correct words per minute)

separately. Another possible application in optometric research is if a number of treatments are to be given to, or tests made on, each of a sample of subjects. In this case, the individual subject is the 'block' and the treatments are given in a random order to each subject.

Model

In the example given in *Table 3*, the fact that subjects are also grouped into age classes gives a more complex model (*Table 1*). In addition to the treatment effect '*a*', this design also includes a term for the age effect '*b*' and the ANOVA table includes an extra term for subject age. In addition to the assumptions made in the randomised design, i.e. homogeneity of variance, additive class effects and normal distribution of errors (Armstrong *et al.*, 2000), this type of design makes the additional assumption that the difference between two treatments is consistent across all replications.

ANOVA

The ANOVA appropriate to the two-way design in randomised blocks is shown in *Table 3*. This design is often used to remove the effect of a particular source of variation from the analysis. For example, if there was significant variation because of the age of the subjects and, if subjects had been allocated to treatments at random, then all the between subject age variation would have been included in the pooled error variance. The effect of this would be to increase the error variance, reduce the 'power' of the experiment and therefore, make it more difficult to demonstrate a possible treatment effect. In a two-way randomised blocks design, however, variation between subjects, attributable to their age, is calculated as a separate effect and therefore, does not appear in the error variance. This may increase the 'power' of the experiment and make it more probable that a treatment effect would be demonstrated.

In the example quoted (*Table 3*), despite the blocking by age, there is no evidence for an effect of coloured filter on reading rates but significant effects were present between 'blocks' presumably reflecting the effect of age.

A comparison of the ANOVA table in *Table 3* with that for a one-way ANOVA in a randomised design (Armstrong *et al.*, 2000) demonstrates that reducing the error variance by blocking has a cost, viz., a reduction in the degrees of freedom (d.f.) of the error variance which makes the estimate of the error variation less reliable. Hence, an experiment in randomised blocks would only be effective if the blocking by age or some other factor reduced the pooled error variance sufficiently to counter the reduction in d.f. It has been estimated that a block-type experiment needs to produce a 14% reduction in the error variance in order to offset the additional unreliability in estimating the pooled error (Cochran and Cox, 1957).

The three-way ANOVA

In the two-way ANOVA in randomised blocks, when treatments are given sequentially to a subject, there is a possible 'carry-over' effect of one treatment on to the next or the subject may become fatigued as the tests proceed. An example of the former might include the sequential application of two drugs without a sufficient recovery period between them and of the latter, reading tests with different filters or magnifiers applied sequentially to the same subject. The solution is to have each combination of treatments given to the same number of subjects such that systematic effects due to treatment order will not create bias in the comparison of the treatment means.

Example

Examples of this type of design are shown in *Table 4*. With two treatments (A and B) and '*n*' subjects, each of the treatment orders AB and BA would be given to *n*/2 subjects. In this case, it has been suggested that the data

Table 4. A three-way ANOVA with different treatments applied in sequence to the same subject

Example 1. Two treatments with '*n*' subjects

Combinations	AB	BA
Subjects	*n*/2	*n*/2

Example 2. Three treatments with '*n*' subjects

Combinations	ABC	ACB	BAC	BCA	CAB	CBA
Subjects	*n*/6	*n*/6	*n*/6	*n*/6	*n*/6	*n*/6

Structure of ANOVA table for three treatments and 36 subjects

Variation	SS	d.f.	MS	*F*
Total		107		
Treatments (a)		2		F_{trts}
Order (c)		5		F_{order}
Subjects (b)		35		F_{subj}
Error		65		

n = number of subjects, trts = treatments, subj = subjects.

could be analysed using '*t*' tests as it may allow for incomplete and non-orthogonal designs (Armitage and Berry, 1987). With three treatments (A, B and C), if all treatment combinations were used, the order of treatments would be ABC, ACB, BAC, BCA, CAB and CBA and each would be given to *n*/6 subjects.

Model

The model for this type of design is given in *Table 1*. In this case, a third term c_k is added which represents the order of the treatment. The term d_{ij} represents the fact that the effect of a treatment may vary from subject to subject while the error term represents measurement errors and non-systematic variation between subjects (Snedecor and Cochran, 1980). Note that this model, as it stands, does not allow for a test of the possible interaction between treatment and order of treatment because this would require a larger factorial-type experiment and a different ANOVA.

ANOVA

In the ANOVA table (*Table 4*), variation attributable to the order of the treatments now appears as a specific experimental effect. This variation will not appear in the pooled error variance or affect the comparison of the treatment means. A limitation of this design, however, is that the number of replications '*n*' must be a multiple of the number of treatments. Hence, with many treatments there will be a large number of possible orders of these treatments and the level of replication will increase accordingly. One method of solving this problem would be to use an incomplete design in which only some of the combinations of treatments would be given. For example, it would be possible to ensure that each treatment was given first, second and third to an equal number of

subjects, e.g. only the combinations ABC, BCA, CAB would be used (Snedecor and Cochran, 1980).

Factorial ANOVA

In a factorial experiment, the effects of a number of different factors can be studied at the same time. Combining factors usually requires fewer experimental subjects or replications than studying each factor individually in a separate experiment. In addition, by contrast with the three-way design, the between treatments or groups sums of squares is partitioned into specific comparisons or 'contrasts' (Ridgman, 1975; Armstrong *et al.*, 2000) which reveal the possible interactive effects between the factors. The interactions between factors often provide the most interesting information from a factorial experiment.

Example

Consider an example of the simplest type of a factorial experiment involving the application of two drugs (A and B) each given at two 'levels' (given or not given) to 24 subjects, randomly allocated, six to each treatment combination (*Table 5*). There are four treatment combinations, no drugs given, either 'A' or 'B' are given separately, or both drugs are given. This type of design is called a 2^2 factorial, i.e. two factors with two levels of each factor.

Model

The model for this design is given in *Table 1*. In this model, x_{ijk} is the value of the '*k*th' replicate of the '*i*th' level of A and the '*j*th' level of B, a_i and b_j are the main effects and $(ab)_{ij}$ represents the two factor interaction between A and B.

Table 5. Factorial ANOVA with two drugs (A and B) given at two levels, given (+) or not given (−) (a 2^2 factorial) with six replications

Design: treatment combinations and orthogonal coefficients	None (1)	+A	+B	+AB
Treatment totals	832	853	881	966
Factorial effects				
A	−1	+1	−1	+1
B	−1	−1	+1	+1
AB	+1	−1	−1	+1
Structure of ANOVA table				
Variation	**SS**	**d.f.**	**MS**	**F**
Total (Drugs)		11 (3)		
A	468	1	468	2.38 ns
B	1094	1	1094	5.57*
AB	171	1	171	0.87 ns
Error	3926	20	196.3	

A, B main effects, AB interaction effect, *p < 0.05, ns = not significant (*p* > 0.05).

ANOVA

As in previous examples, the total sums of squares can be broken down into that associated with differences between the effects of the drugs and error (*Table 5*). In this case, the between treatments sums of squares can be broken down further into 'contrasts' which describe the main effects of 'A' and 'B' and the interaction effect 'A × B'. These effects are linear combinations of the means, each being multiplied by a number or 'coefficient' to calculate a particular effect. In fact, the meaning of an effect can often be appreciated by studying these coefficients (*Table 5*). The main effect of drug 'A' is calculated from those groups of subjects which receive drug 'A' (+) compared with those which do not (−). Note that in a factorial design, every observation is used in the estimate of the effect of every factor. Hence, factorial designs have 'internal replication' and this may be an important consideration in deciding the number of subjects to use. The main effect of 'B' is calculated similarly to that of 'A'. By contrast, the two–factor interaction ('A × B') can be interpreted as a comparison of the effect of the combined action of 'A' and 'B' with the individual effects of 'A' and 'B'. A significant interaction term would imply that the effects of 'A' and 'B' were not additive, i.e. the effect of the combination 'AB' would not be predictable from knowing the individual effects of 'A' and 'B'. In the quoted example, there is no significant effect of drug A, the effect of drug B is significant and the non-significant interaction indicates that the effect of B was consistent regardless of whether A was given or not.

Note that in a 2^2 factorial, partitioning the treatments sums of squares into factorial effects provides all the information necessary for interpreting the results of the experiment and further '*posthoc*' tests would not be needed. However, with more complex factorial designs, e.g. those with more than two levels of each factor, further tests may be required to interpret the main effect or an interaction. With factorial designs, it is better to define specific comparisons before the experiment is carried out rather than to rely on '*posthoc*' tests. Factorial experiments can be carried out in a completely randomised design, in randomised blocks or in a more complex design. The relative advantages of these designs are the same as for the one-way design.

Factorial ANOVA, split-plot design

In the 2^2 factorial described above, the experimental subjects were assigned at random to all possible combinations of the two factors. However, in some designs, the two factors are not equivalent to each other. A common case, called a split-plot design, arises when one factor can be considered to be a major factor and the other a minor factor.

Example

In optometry research, this situation could arise if measurements were made on both the right and left eyes of subjects employing two different subject groups (*Table 6*). As an example, an investigator wished to study whether IOP in right and left eyes was elevated in patients with high blood pressure. Alternatively, a measurement may be made on subjects within two treatment groups employing different methods or equipment. The problem that arises in these types of design is the dependence or correlation between the measurements made on a subject (Rosner, 1982). In these experiments, the subject group would be the major factor while 'which eye' or 'which method' would be regarded as a minor factor. The difference between this and an ordinary factorial design is that previously, all treatment combinations were assigned at random to replicates whereas in a split-plot design, the subplot

Table 6. Factorial ANOVA, split-plot design with two subject groups (normal and elevated blood pressure) with three subjects (P1, P2, P3) in each group and the left (L) and right (R) eye studied from each patient. Data are intraocular pressures

Design	P1	P2	P3	
Control				
R eye	17.3	16.9	14.7	
L eye	17.1	16.5	14.3	
ARMD				
R eye	21.4	24.3	21.4	
L eye	20.7	22.1	24.2	
ANOVA table:				
Variation	**SS**	**d.f.**	**MS**	***F***
Subject group	115.94	1	115.94	34.4**
Main plot error	13.48	4	3.37	
Right/Left eye	0.10	1	0.10	0.06 ns
Interaction	0.067	1	0.067	0.04 ns
Sub plot error	6.59	4	1.65	

**$p < 0.01$; ns = not significant.

treatments are randomised over the whole large block but only within each main plot. In some applications, experimenters may subdivide the subplots further to give a split-split-plot design (Snedecor and Cochran, 1980).

Model

The model for a two factor split-plot design is given in *Table 1*. In this case, M represents main-plot treatments, B blocks and, T subplot treatments. The symbols '*i*' and '*j*' indicate the main-plots while '*k*' identifies the subplot within the main-plot. The two components of error e_{ij} and d_{ijk} represent the fact that the error variation between main-plots is likely to be different from that between subplots. For example, one might expect there to be less natural variation when comparing eyes within a subject (subplot) than the eyes of different subjects (main-plot). An alternative model suggested by Rosner (1982) would be to consider the x_{ij} as distributed about the same mean and variance but for any two members of the same class to be correlated, the degree of correlation being measured by the intra-class correlation coefficient (Snedecor and Cochran, 1980).

ANOVA

The resulting ANOVA (*Table 6*) is more complex than that of a simple factorial design because of the different error terms. Hence, in a two-factor, split-plot ANOVA, two errors are calculated, the main-plot error is used to test the main effect of subject group while the subplot error is used to test the main effect of eyes and the possible interaction between the factors. In the quoted example, there is a significant increase in IOP in patients with elevated blood pressure but no difference between eyes; the non-significant interaction suggesting that the elevation in IOP was consistent in both eyes. The subplot error is usually smaller that the main-plot error and also has more d.f. Hence, such an experimental design will usually estimate the main effect of the subplot factor and its interaction more accurately than the main effect of the major factor. Some experimenters will deliberately design an experiment as a 'split-plot' to take advantage of this property.

A disadvantage of such a design, however, is that occasionally, the main effect of the major factor may be large but not significant while the main effect of the minor factor and its interaction may be significant but too small to be important. In addition, a common mistake is for researchers to analyse a split-plot design as if it were a fully randomised two-factor experiment. In this case, the single pooled error variance will either be too small or too large for testing the individual treatment effects and the wrong conclusions could be drawn from the experiment. To decide whether a particular experiment is a split-plot it is useful to

consider the following: (1) Are the factors equivalent or does one appear to be subordinate to the other? (2) Is there any restriction in how replicates were assigned to the treatment combinations? and (3) Is the error variation likely to be the same for each factor?

Caution should also be employed in the use of '*posthoc*' tests in the case of a split-plot design. '*Posthoc*' tests assume that the observations taken on a given subject are uncorrelated so that the subplot factor group means are not related. This is not likely to be the case in split-plot experiment because some correlation between measurements made on the right and left eye is inevitable. Standard errors appropriate to the split-plot design can be calculated (Cochran and Cox, 1957; Freese, 1984) and can be used, with caution, to make specific comparisons between the treatment means. However, a better method is to partition the sums of squares associated with main effects and interaction into specific contrasts and to test each against the appropriate error (Snedecor and Cochran, 1980).

Factorial ANOVA, repeated measures design

The repeated measures factorial design is a special case of the split-plot type experiment in which measurements on the experimental subjects are made sequentially over several intervals of time. The ANOVA is identical to the preceeding example but with time constituting the subplot factor. Repeated measurements made on a single individual are likely to be highly correlated and therefore the usual '*posthoc*' tests cannot be used. Nevertheless, it is possible to partition the main effects and interaction sums of squares into contrasts. In a repeated measures design, the shape of the reponse curve, i.e. the regression of the measured variable on time, may be of particular interest. A significant interaction between the main-plot factor and time would indicate that the response curve with time varied at different levels of the main-plot factor.

Non-conformity to model

In the ANOVAs described above, the model specifies that the effects of different factors should be additive and that the errors are normally distributed with the same variance. It is unlikely that these conditions are ever realised in real experiments and it is reasonable to consider the consequences of failure of these assumptions.

A major error in the measurement of the data can distort the mean of a treatment considerably and, by increasing the error variance, can have a profound effect on the experiment as a whole. An investigation into 'suspicious' values should always be made. Such a process may reveal a mistake and enable the correct value to be substituted or if not, rejecting the aberrant

observation and analyzing the data without it. A number of rules for rejecting such observations are given by Snedecor and Cochran (1980).

Lack of independence of the errors can arise if a faulty experimental design is employed. For example, if all replications of a given treatment were processed at the same time by the same person or by the same machine; different technicians or machines being employed for the other treatments, then there will be positive correlations between replicates within treatments. The most effective precaution against this type of effect is to use appropriate randomization at every stage of the experiment. If different treatments have different error variances (heterogenous variances), the *F*-test may indicate a significant result when no true effects are present. One method of solving this problem is to omit those treatments which are substantially different from the rest. In addition, if within the different treatments, the standard deviation is proportional to the mean, such that effects are proportional rather than additive, a logarithmic transformation may stabilise the variance (Armstrong *et al.*, 2000).

Non-normal distribution of the errors will also tend to produce too many significant results in an ANOVA. The solution may be to transform the data, e.g. counts of rare events may be distributed according to the Poisson distribution and a transformation to square roots may be appropriate. In addition, if the data are proportions or percentages, an angular transformation may be necessary (Snedecor and Cochran, 1980).

Number of replications in an experiment

In our previous article (Armstrong *et al.*, 2000), we described a simple method of obtaining an approximate estimate of the number of replications required in an experiment to have a good chance of revealing a particular effect. Such an approach may provide a reasonable approximation in simple experiments, such as the one-way design, but are unlikely to be appropriate in more complex experiments with many treatments or factorial combinations. In more complex experiments, it is better to consider the number of d.f. of the error term rather than the number of replications per treatment group. The number of d.f. depends on both the number of treatments and the number of replicates. It is not possible to define the lower limit of the d.f. for all experiments, but a good standard would be to try and achieve at least 15 d.f. for the error term if possible (Ridgman, 1975). Hence, in an experiment with four treatments, five replicates per treatment would be required to provide approximately 15 d.f. for the error term. In more complex designs, such as the split-plot or repeated measures design, which usually have more precision in the estimation of the

subplot factor and interaction, then it would be appropriate to have at least 15 d.f. for the subplot error.

Conclusions

The key to the correct application of ANOVA in optometric research is careful experimental design and matching the correct analysis to that design. The following points should therefore, be considered before designing any experiment:

(1) In a single factor design, ensure that the factor is identified as a 'fixed' or 'random effect' factor.
(2) In more complex designs, with more than one factor, there may be a mixture of fixed and random effect factors present so ensure that each factor is clearly identified.
(3) Where replicates can be grouped or blocked, the advantages of a randomised blocks design should be considered. There should be evidence, however, that blocking can sufficiently reduce the error variation (by at least 14%) to counter the loss of d.f. compared with a randomised design.
(4) Where different treatments are applied sequentially to a patient, the advantages of a three-way design in which the different orders of the treatments are included as an 'effect' should be considered.
(5) If treatments can be expressed as factors then combining different factors to make a more efficient experiment and to measure possible factor interactions should always be considered.
(6) The effect of 'internal replication' should be taken into account in a factorial design in deciding the number of replications to be used. Where possible, each error term of the ANOVA should have at least 15 d.f.
(7) Consider carefully whether a particular factorial design can be considered to be a split-plot or a repeated measures design. If such a design is appropriate, consider how to continue the analysis bearing in mind the problem of using *post hoc* tests in this situation.
(8) Data should be checked for gross errors, lack of independence of the errors, non-additivity and non-normality. In some cases, transformation of the data may be necessary before the analysis is carried out.
(9) If there is any doubt about the above issues, the researcher should seek advice from a statistician with experience of optometric research before carrying out the experiment. It is particularly important to check for model assumptions and that the d.f. and error terms are correct. An erroneous design will not test a null hypothesis adequately and once committed to such a design, there may be little a statistician can do to help.

Editor's Note: This article and the first two discussions were presented orally at the *Technometrics* session of the 36th Annual Fall Technical Conference held in Philadelphia, Pennsylvania, October 8–9, 1992. The conference was cosponsored by the Chemical and Process Industries, the Statistics Divisions of the American Society for Quality Control, and the Section on Physical and Engineering Sciences of the American Statistical Association.

A Systematic Approach to Planning for a Designed Industrial Experiment

David E. Coleman

Alcoa Laboratories
Alcoa Center, PA   15069

Douglas C. Montgomery

Industrial Engineering Department
Arizona State University
Tempe, AZ   85287

Design of experiments and analysis of data from designed experiments are well-established methodologies in which statisticians are formally trained. Another critical and rarely taught skill is the planning that precedes designing an experiment. This article suggests a set of tools for presenting generic technical issues and experimental features found in industrial experiments. These tools are predesign experiment guide sheets to systematize the planning process and to produce organized written documentation. They also help experimenters discuss complex trade-offs between practical limitations and statistical preferences in the experiment. A case study involving the (computer numerical control) CNC-machining of jet engine impellers is included.

KEY WORDS: Industrial experimental design; Measurement error; Nuisance factors; Statistical consulting.

1. INTRODUCTION

1.1 A Consulting Scenario

Consider the following scenario: An experimenter from the process engineering group comes to you and says: "We are manufacturing impellers that are used in a jet turbine engine. To achieve the claimed performance objectives, we must produce parts with blade profiles that closely match the engineering design requirements. I want to study the effect of different tool vendors and machine set-up parameters on the dimensional variability of the parts produced on the machines in our CNC-machine center."

Many experimental design applications in industry begin with such a statement. It is well recognized that the planning activities that precede the actual experiment are critical to successful solution of the experimenters' problem (e.g., see Box, Hunter, and Hunter 1978; Hahn 1977, 1984; Montgomery 1991; Natrella 1979). Montgomery (1991) presented a seven-step approach for planning experiments, summarized in Table 1. The first three of these steps constitutes the preexperiment planning phase. The detailed, specific activities in this phase are the focus of this article. The emphasis is planning for a screening ex-

periment, or a step in sequential experimentation on an existing product/process, off-line or on-line. Many of the issues addressed, however, also apply to new products/processes or research and development (R&D) and to various additional experimental goals, such as optimization and robustness studies.

1.2 A Gap

It is often said that no experiment goes exactly as planned, and this is true of most industrial experiments. Why? One reason is that statisticians who design experiments with scientists and engineers (the "experimenters") usually have to bridge a gap in knowledge and experience. The consequences of not bridging this gap can be serious.

The statistician's lack of domain knowledge can lead to:

1. Unwarranted assumptions of process stability during experimentation
2. Undesirable combinations of control-variable levels in the design
3. Violation or lack of exploitation of known physical laws
4. Unreasonably large or small designs

1. Recognition of and statement of the problem
2.* Choice of factors and levels
3.* Selection of the response variable(s)
4. Choice of experimental design
5. Conduction of the experiment
6. Data analysis
7. Conclusions and recommendations

*In some situations, steps 2 and 3 can be reversed.

5. Inappropriate confounding

6. Inadequate measurement precision of responses or factors

7. Unacceptable prediction error

8. Undesirable run order

The experimenter's lack of statistical knowledge can lead to:

1. Inappropriate control-variable settings (e.g., range too small to observe an effect or range so large that irrelevant mechanisms drive the response variable)

2. Misunderstanding of the nature of interaction effects, resulting in unwisely confounded designs

3. Experimental design or results corrupted by measurement error or setting error

4. Inadequate identification of factors to be "held constant" or treated as nuisance factors, causing distorted results

5. Misinterpretation of past experiment results, affecting selection of response variables or control variables and their ranges

6. Lack of appreciation of different levels of experimental error, leading to incorrect tests of significance

This article attempts to help bridge the gap by providing a systematic framework for predesign information gathering and planning. Specifically, we present *guide sheets* to direct this effort. The use of the guide sheets is illustrated through the (computer numerical control) CNC-machinery example briefly presented previously. This article is a consolidation and extension of the discussion by Hahn (1984), Box et al. (1978), Montgomery (1991), Natrella (1979), Bishop, Petersen, and Traysen (1982), and Hoadley and Kettenring (1990).

The guide sheets are designed to be discussed and filled out by a multidisciplinary *experimentation team* consisting of engineers, scientists, technicians/operators, managers, and process experts. These sheets are particularly appropriate for complex experiments and for people with limited experience in designing experiments.

The sheets are intended to encourage the discussion and resolution of generic technical issues needed *before* the experimental design is developed. Hahn

(1984) listed most of these issues and made the recommendation, "The major mode of communication between the experimenter and the statistician should be face-to-face discussion. The experimenter should, however, also be encouraged to document as much of the above information as possible ahead of time" (p. 25). Unfortunately, as Hahn observed, "Not all experimenters are willing to prepare initial documentation" (p. 26). Moreover, not all of the relevant issues may be thoroughly thought out—hence, the need for face-to-face discussions, during which, as Hahn advised, "The statistician's major functions are to help structure the problem, to identify important issues and practical constraints, and to indicate the effect of various compromises on the inferences that can be validly drawn for the experimental data" (p. 21).

The guide sheets proposed in this article outline a systematic "script" for the verbal interaction among the people on the experimentation team. When the guide sheets are completed, the team should be well equipped to proceed with the task of designing the experiment, taking into account the needs and constraints thus identified.

2. PREDESIGN MASTER GUIDE SHEET AND SUPPLEMENTARY SHEETS

The guide sheets consist of a "Master Guide Sheet," plus supplementary sheets and two tutorials. These are schematically illustrated in Figure 1. The supplementary sheets are often more convenient for items 3–7.

The Master Guide Sheet is shown in Figure 2. It is stripped of the blank space usually provided to fill in the information. Blank copies will be provided by the authors on request.

Discussion of issues related to different pieces of the Master Guide Sheet and the supplementary sheets follows.

Writing the objective (item 2, Fig. 2) is harder than it appears to most experimenters. Objectives should be (a) unbiased, (b) specific, (c) measurable, and (d) of practical consequence. To be *unbiased*, the experimentation team must encourage participation by knowledgeable and interested people with diverse perspectives. The data will be allowed to speak for themselves. To be *specific and measurable*, the objectives should be detailed and stated so that it is clear whether they have been met. To be of *practical consequence*, there should be something that will be done differently as a result of the outcome of the experiment. This might be a change in R&D direction, a change in process, or a new experiment. Conducting an experiment constitutes an expenditure of resources *for some purpose*.

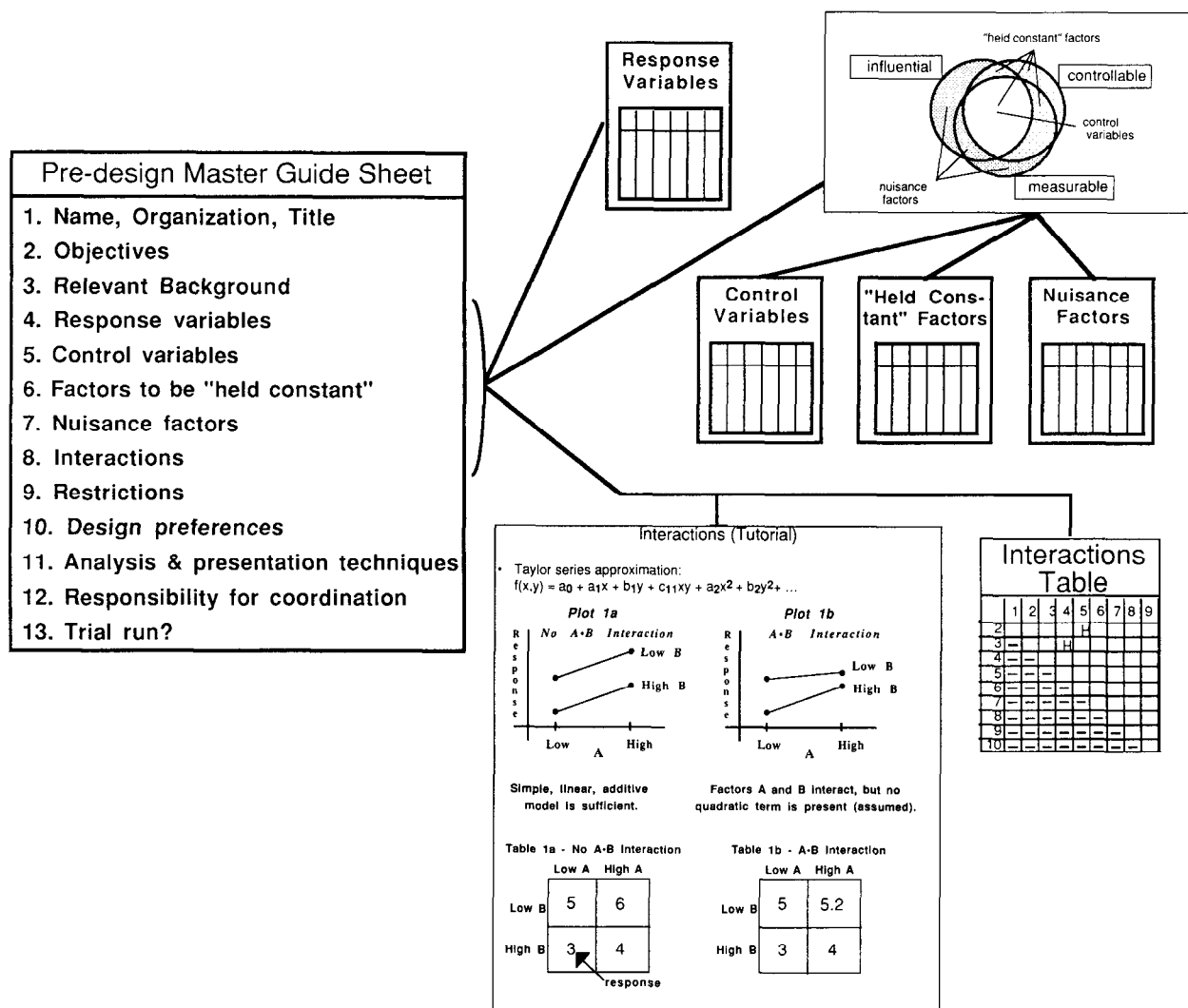Thus experimental objectives should not be stated

Pre-design Master Guide Sheet
1. Name, Organization, Title
2. Objectives
3. Relevant Background
4. Response variables
5. Control variables
6. Factors to be "held constant"
7. Nuisance factors
8. Interactions
9. Restrictions
10. Design preferences
11. Analysis & presentation techniques
12. Responsibility for coordination
13. Trial run?

Response Variables

"held constant" factors

influential

controllable

control variables

nuisance factors

measurable

Control Variables

"Held Constant" Factors

Nuisance Factors

Interactions (Tutorial)

- Taylor series approximation:

$f(x,y) = a_0 + a_1 x + b_1 y + c_{11} xy + a_2 x^2 + b_2 y^2 + \ldots$

Plot 1a

Response

No A·B Interaction

Low B

High B

Low   High
        A

Plot 1b

Response

A·B Interaction

Low B

High B

Low   High
        A

Simple, linear, additive model is sufficient.

Factors A and B interact, but no quadratic term is present (assumed).

Table 1a - No A·B Interaction

	Low A	High A
Low B	5	6
High B	3	4

response

Table 1b - A·B Interaction

	Low A	High A
Low B	5	5.2
High B	3	4

Interactions Table

	1	2	3	4	5	6	7	8	9
2	—								
3	—		H						
4	—	—							
5	—	—	—						
6	—	—	—	—					
7	—	—	—	—	—				
8	—	—	—	—	—	—			
9	—	—	—	—	—	—	—		
10	—	—	—	—	—	—	—	—	

Figure 1. Structure of Predesign Experiment Guide Sheets.

as, "To show that catalyst z14 works better than catalyst d12, if the technician adjusts the electrode voltage just right." A better objective would be: "To quantify the efficiency difference, Δ, between catalysts z14 and d12 for electrode voltages 7, 8, and 9 in the ABC conversion process—and assess statistical significance (compare to 95%) and practical significance ($\Delta > 3\%$), perhaps economically justifying one catalyst over the other."

As Box et al. (1978, p. 15) put it (paraphrased), the statistician or other members of the experimentation team should "ensure that all interested parties agree on the objectives, agree on what criteria will determine that the objectives have been reached, and arrange that, if the objectives change, all interested parties will be made aware of that fact and will agree on the new objectives and criteria." Even experimenters in the physical sciences—who have been trained in the scientific method—sometimes need prodding in this.

The objective of the experiment can be met if the predesign planning is thorough, an appropriate design is selected, the experiment is successfully conducted, the data are analyzed correctly, and the results are effectively reported. By using a systematic approach to predesign planning, there is greater likelihood that the first three conditions will occur. This increases the likelihood of the fourth. Then the experiment is likely to produce its primary product— new knowledge.

2.1 Relevant Background

The relevant background supporting the objectives should include information from previous experiments, routinely collected observational data, physical laws, and expert opinion. The purposes of providing such information are (a) to establish a context for the experiment to clearly understand what new knowledge can be gained; (b) to motivate discussion about the relevant domain knowledge, since

1. **Experimenter's Name and Organization:**   **Brief Title of Experiment:**
2. **Objectives of the experiment** (should be unbiased, specific, measurable, and of practical consequence):
3. **Relevant background** on response and control variables: (a) theoretical relationships; (b) expert knowledge/experience; (c) previous experiments. Where does this experiment fit into the study of the process or system?:
4. List: (a) each **response variable**, (b) the normal response variable level at which the process runs, the distribution or range of normal operation, (c) the precision or range to which it can be measured (and how):
5. List: (a) each **control variable**, (b) the normal control variable level at which the process is run, and the distribution or range of normal operation, (c) the precision (s) or range to which it can be set (for the experiment, not ordinary plant operations) and the precision to which it can be measured, (d) the proposed control variable settings, and (e) the predicted effect (at least qualitative) that the settings will have on each response variable:
6. List: (a) each factor to be **"held constant"** in the experiment, (b) its desired level and allowable s or range of variation, (c) the precision or range to which it can measured (and how), (d) how it can be controlled, and (e) its expected impact, if any, on each of the responses:
7. List: (a) each **nuisance factor** (perhaps time-varying), (b) measurement precision, (c) strategy (e.g., blocking, randomization, or selection), and (d) anticipated effect:
8. List and label known or suspected **interactions:**
9. List **restrictions** on the experiment, e.g., ease of changing control variables, methods of data acquisition, materials, duration, number of runs, type of experimental unit (need for a split-plot design), "illegal" or irrelevant experimental regions, limits to randomization, run order, cost of changing a control variable setting, etc.:
10. Give current **design preferences**, if any, and reasons for preference, including blocking and randomization:
11. If possible, propose **analysis and presentation techniques**, e.g., plots, ANOVA, regression, plots, t-tests, etc.:
12. **Who will be responsible for the coordination of the experiment?**
13. **Should trial runs be conducted?   Why / why not?**

Figure 2. Predesign Master Guide Sheet. This guide can be used to help plan and design an experiment. It serves as a checklist to accelerate experimentation and ensures that results are not corrupted for lack of careful planning. Note that it may not be possible to answer all questions completely. If convenient, use the supplementary sheets for 4–8.

such discussion may change the consensus of the group, hence the experiment; and (c) to uncover possible experimental regions of particular interest and others that should be avoided. With this background, we reduce the risks of naive empiricism and duplication of effort.

For the CNC-matching problem introduced earlier, we have the guide sheet shown in Figure 3.

3. RESPONSE VARIABLES

As mentioned previously, items 4–8 on the guide sheet are most conveniently handled using the supplementary sheets. The first one is for response variables, as shown in Table 2.

Response variables come to mind easily for most experimenters, at least superficially; they know what outcomes they want to change—a strength, a failure

rate, a concentration, or a yield. What makes a good response variable? The answer to this question is complex. A complete answer is beyond the scope of this article, but some guidelines can be given. A response variable

1. Is preferably a continuous variable. Typically, this will be a variable that reflects the continuum of a physical property, such as weight, temperature, voltage, length, or concentration. Binary and ordinal variables have much less information content—much as the raw values are more informative than histograms that have wide bins. Note that being continuous *with respect to a control variable* may be important. If a response variable has, perhaps, a steep sigmoidal response to a control variable, it is effectively binary as that variable changes. For example,

1.Experimenter's Name and Organization: John Smith, Process Eng. Group Brief Title of Experiment: CNC Machining Study
2. Objectives of the experiment (should be unbiased, specific, measurable, and of practical consequence):
For machined titanium forgings, quantify the effects of tool vendor; shifts in a-axis, x-axis, y-axis, and z-axis; spindle speed; fixture height; feed rate; and spindle position on the average and variability in blade profile for class X impellers, such as shown in Figure 4.
3. Relevant background on response and control variables: (a) theoretical relationships; (b) expert knowledge/experience; (c) previous experiments. Where does this experiment fit into the study of the process or system?: (a) Because of tool geometry, x-axis shifts would be expected to produce thinner blades, an undesirable characteristic of the airfoil. (b) This family of parts has been produced for over 10 years; historical experience indicates that externally reground tools do not perform as well as those from the "internal" vendor (our own regrind operation). (c) Smith (1987) observed in an internal process engineering study that current spindle speeds and feed rates work well in producing parts that are at the nominal profile required by the engineering drawings - but no study was done of the sensitivity to variations in set-up parameters.
Results of this experiment will be used to determine machine set-up parameters for impeller machining. A robust process is desirable; that is, on-target and low variability performance regardless of which tool vendor is used.

Figure 3. Beginning of Guide Sheet for CNC-Machining Study.

weight of precipitate as a function of catalyst may be near zero for the selected low levels of catalyst and near maximum for the high levels.

2. Should capture, as much as possible, a quantity or quality of interest for the experimental unit. For example, if the experimental unit is an ingot and a response is T = temperature, it may matter whether T is taken at a single point or averaged over a surface region, the entire surface area, or the entire ingot volume.

3. Should be in appropriate units. The units may

be absolute, such as pounds, degrees centigrade, or meters. They may be relative units, such as percent of concentration by weight or by volume or proportional deviation from a standard. What is "appropriate" may be determined by an empirical or first-principles model, such as using absolute units in $E = mc^2$, or it may be determined by practical limitations, such as using percent of concentration by weight because the experimental samples are not all the same weight.

4. Should be associated with a target or desirable

Table 2. Response Variables

Response variable (units)	Normal operating level and range	Measurement precision, accuracy— how known?	Relationship of response variable to objective
Blade profile (inches)	Nominal (target) ± 1 × 10⁻³ inches to ± 2 × 10⁻³ inches at all points	$\sigma_E \approx 1 \times 10^{-5}$ inches from a coordinate measurement machine capability study	Estimate mean absolute difference from target and standard deviation of difference
Surface finish	Smooth to rough (requiring hand finish)	Visual criterion (compare to standards)	Should be as smooth as possible
Surface defect count	Typically 0 to 10	Visual criterion (compare to standards)	Must not be excessive in number or magnitude

condition (which motivates the experiment). Such a comparison might be used to derive "performance measures" from response-variable outcomes. For example, with CNC-machining, blade profile is a response variable, and it is compared to the target profile by computing differences at certain locations. Mean absolute difference and the standard deviation of the differences are performance measures for the various experimental conditions. They can be analyzed separately or by using the standard deviations to compute weights for the mean analysis.

5. Is preferably obtained by nondestructive and nondamaging methods so that repeated measures can be made and measurement error can be quantified.

6. Should not be near a natural boundary. Otherwise, the variable will not discriminate well. For example, it is hard to distinguish a yield of 99.5% from 99.8%, and it is hard to detect and distinguish contamination levels near 0.

7. Preferably has constant variance over the range of experimentation.

There are other important characteristics of response variables that the experimenters may not have considered or communicated to the whole experimentation team. This sheet helps to draw them out: (a) current use, if any (col. 2); (b) ability to measure (col. 3); and (c) the knowledge sought through experimentation (col. 4).

It is helpful to know the current state of use, and if it is unknown, the experimenters are advised to include some trial runs prior to the experiment or "checkpoint" runs during the experiment (perhaps these data have not been previously acquired). The current distribution serves as one of several possible reference distributions for judging the *practical* magnitude of the effects observed. Given a typical standard deviation for a response variable of σ, a low-to-high control-variable effect of $\sigma/2$ may be of no practical significance, but one of 4σ may be important. Another advantage to knowing the current state of use is a check on credibility. Process or design limitations may constrain a response variable to be bounded on one or two sides. An experimental result outside that range may be erroneous or due to an abnormal mechanism (which may, however, be of interest).

Measurement precision (and, in some cases, bias) and how to obtain it (i.e., choice of measurement system or repeated measurements) is a thorn in the flesh for many experimenters. The admonition of Eisenhart (1962) serves as a relevant (if overstated) warning, "until a measurement operation . . . has attained a state of statistical control it cannot be regarded in any logical sense as measuring anything at all" (p. 162). It has been our experience that many

experimenters do not know the state of control nor the precision and bias of most measurement systems measuring a response or a control variable. The measurement systems were not *useless*, they were just of unknown *utility*. Important and poorly understood systems should be evaluated with a measurement capability study, a designed experiment of the measurement process. As a compromise, one might be forced to resort to historical data and experience or weaken the experimental objective to obtain ranking, selection, or a binary response instead of quantification.

The relationship of a response variable to the objective may be direct. An objective may be defined in terms of a response variable—for example, "to quantify the effect that thermal cycle B has on tensile strength measured on customer qualifying tester X." In the case of CNC-machining, a response variable is blade profile (see Fig. 4). This is related to the objective through two measurement-performance indicators—mean absolute difference of blade profile and the target, and standard deviation of the difference. Sometimes a response variable may be a *surrogate* for the true response of interest. This is often the case in destructive testing, in which a standard stress-to-fracture test, for example, represents performance under conditions of use. Another example is yield rate or failure rate, which are inferior responses that often represent where a specification falls relative to a distribution of continuous-scale values (the collection of which provides superior information).

As discussed previously, the relationship of a response variable to the objective may be through *performance measures* that involve a comparison of the response to a target or desirable outcome.

4. CONTROL VARIABLES

As with response variables, most investigators can easily generate a list of candidate control variables.

Control variables can be attribute or continuous. They can be narrowly defined, such as "percent of copper, by weight," or broadly defined, such as "comparably equipped pc: Apple or IBM." In either case, control variables should be explicitly defined.

When discussing potential control variables with experimenters, it may be helpful to anticipate that *held-constant factors* and *nuisance factors* must also be identified. Figure 5 is a Venn diagram that can be used to help select and prioritize candidate factors. It illustrates different categories of factors that affect response variables, based on three key characteristics—magnitude of influence on response variables, degree of controllability, and measurability (e.g., precision and accuracy). Each type of factor is discussed in detail in following sections. A descrip-
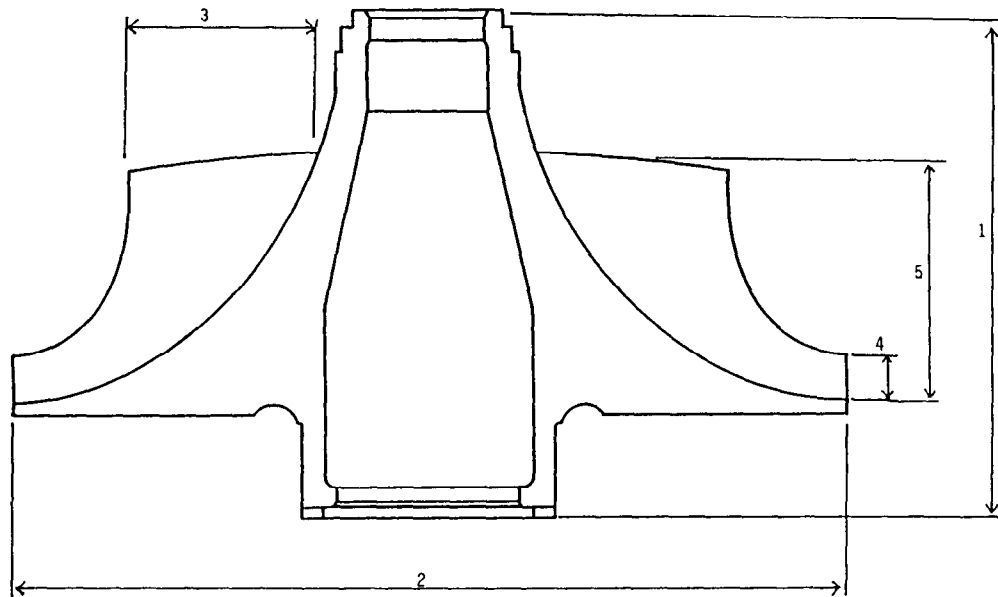
Figure 4. Jet Engine Impeller (side view; z axis is vertical, x axis is horizontal, and y axis is into the page): 1. Height of Wheel; 2. Diameter of Wheel; 3. Inducer Blade Height; 4. Exducer Blade Height; 5. Z Height of Blade.

tion of the diagram is as follows:

1. Control variables are measurable, controllable, and thought to be (very) influential.

2. Held-constant factors are controlled.

3. Nuisance factors are uncontrolled factors (either they cannot be controlled, or they are allowed to vary).

In discussing different variables and factors the team may choose to reassign variables from one group to another, and this is part of the ordinary process for planning a designed experiment. For the CNC-machining problem, the control-variable information was developed as shown in Table 3; those below the space are considered to be of secondary importance.

Similar to the response variables sheet, the control variables sheet solicits information about (a) current use (col. 2), (b) ability to measure and set (col. 3), and (c) knowledge sought through experimentation (cols. 4–5).
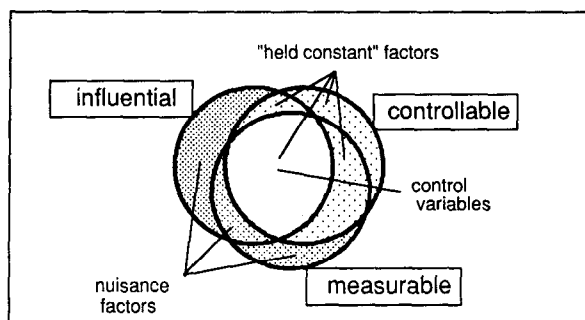

Figure 5. Different Categories of Factors Affecting Response Variables.

4.1 Current Use (col. 2)

There are two reasons it helps to know the allowed ranges and nominal values of control variables under current use. First, the degree to which historical process data can be used to gain relevant knowledge may be revealed. This is discussed in Section 4.2. Second, the experimenter should select a range large enough to produce an observable effect and to span a good proportion of the operating range, yet not choose so great a range that no empirical model can be postulated for the region, as discussed in Section 4.3. In some, less mature experimental situations, there may be no well-defined "current use," in which case trial runs before or during experimentation are helpful—as they are with response variables.

4.2 Ability to Measure and Set (col. 3)

With control variables, there is an additional consideration rarely mentioned in the literature. The experimentation team not only needs to know how *measurements will be obtained* and the precision of measurement, σ_m, but also *how the control variable settings will be obtained* and "setting error," ε_s. These different types of deviation from the ideal have different effects on experimentation. Large σ_m will mean that either errors-in-variables methods will have to be used (e.g., methods that will allow estimation of bias in effects estimates) or, alternatively, many samples will have to be collected for measurement during experimentation to get an acceptably small $\sigma_m/\sqrt{n}$, especially if $|\varepsilon_s|$ is also large. If $|\varepsilon_s|$ is large, traditional, class-variable-based analysis of variance will have to

Table 3. Control Variables

Control variable (units)	Normal level and range	Measurement precision and setting error— how known?	Proposed settings, based on predicted effects	Predicted effects (for various responses)
x-axis shift* (inches)	0—.020 inches	.001 inches (experience)	0, .015 inches	Difference ↗
y-axis shift* (inches)	0—.020 inches	.001 inches (experience)	0, .015 inches	Difference ↗
z-axis shift* (inches)	0—.020 inches	.001 inches (experience)	?	Difference ↗
Tool vendor	Internal, external	—	Internal, external	External is more variable
a-axis shift* (degrees)	0—.030 degrees	.001 degrees (guess)	0, .030 degrees	Unknown
Spindle speed (% of nominal)	85—115%	~1% (indicator on control panel)	90%, 110%	None?
Fixture height	0—.025 inches	.002 inches (guess)	0, .015 inches	Unknown
Feed rate (% of nominal)	90—110%	~1% (indicator on control panel)	90%, 110%	None?

*The x, y, and z axes are used to refer to the part *and* the CNC machine. The a axis refers only to the machine.

be replaced by regression analysis. The result of large setting variation may be unwanted aliasing, greater prediction error, violation of experiment constraints, and difficulty conducting split-plot analyses.

Often, one finds that $\sigma_m \approx |\varepsilon_s|$, such as when the measurement system is part of a controller, and equilibrium conditions can be achieved. Measurement precision and setting error are not always comparable, however. For example, $\sigma_m < |\varepsilon_s|$ is not unusual for a continuous-batch mixing process. Suppose that the concentration of constituent A is at 10% and is continuously reduced towards a target of 5%. Batches might be produced with concentrations of 10%, 7%, and 4%. In this case, perhaps $|\varepsilon_s| \approx 1\%$, but a spectrograph may measure with $\sigma_m \approx .1\%$. Another example is a thermostat, which often provides $\sigma_m < |\varepsilon_s|$, especially if it has a "dead zone" in its logic.

Alternatively, one may find $\sigma_m > |\varepsilon_s|$. For example, physical laws may make it possible to accurately *set* gas pressure in a sealed cavity by setting gas temperature, but there may be no precise way to directly *measure* pressure.

4.3 Knowledge Sought Through Experimentation

In the design of experiments classes he teaches at Alcoa, J. S. Hunter gives a rule of thumb for experiments on existing processes. For each control variable, low/high settings should be selected to cause a predicted effect (main effect) for the "key" response variable equal to one standard deviation of its variation in ordinary use, σ_p (if there is "ordinary use"). This is a large enough change in response to have practical consequence and also large enough to likely be detected if measurement error is negligible and the experiment has enough runs. If the rule of thumb is followed, every control variable has "equal opportunity" to affect the response variable.

Naturally, it is harder to suggest such a rule for immature processes. Moreover, other issues and constraints must be taken into account when settings are selected—safety, discreteness of settings, process constraints, ease of changing a setting, and so forth. These are solicited in item 8 of the guide sheet.

Predicted effects for the response variables may be available from the knowledge sources previously listed—theory, experts, and experiments. Quantitative predicted effects are preferable, but experimenters may not be able to provide more than qualitative indications. Even if uncertain, the exercise of attempting to predict the outcome of the experiment before it is run can foster good interaction within the experimentation team and often leads to revised choices of settings. An additional advantage is that the predictions will always be wrong, so it is easier to see what knowledge has been gained through experimentation.

5. HELD-CONSTANT FACTORS

Held-constant factors are controllable factors whose effects are not of interest in this experiment. Most experimenters think in these terms: "For this experiment, I want to study the effect of factors A, B, and C on responses y_1 and y_2, but all other control variables should be held at their nominal settings, and I do not want extraneous factors distorting the results." This sheet was developed to ensure the "all other control variables held at their nominal settings" condition. (The next sheet is used to help ensure that "there are no extraneous factors distorting the results.") For the CNC-machining example, the held-constant factors are as shown in Table 4.

The sheet in Table 4 can force helpful discussion about which factors are adequately controlled and which factors are not. In so doing, it is often necessary to consult experts to help prioritize factors, recommend preexperiment studies to assess control, or develop control strategy.

For example, in the CNC-machining case, this sheet resulted in the recognition of the fact that the machine had to be fully warmed up before cutting any blade forgings. The actual procedure used was to mount the forged blanks on the machine spindles and run a 30-minute cycle without the cutting tool engaged. This would allow all machine parts and the lubricant to reach normal, steady-state operating temperature. The use of a "typical" (i.e., "mid-level") operator and the blocking of the blank forgings by lot number were decisions made for experimental insurance, although neither variable was expected to have important effects. Not that it is not practical or desirable to hold some factors constant. For example, although it might be ideal to have experimental material from only one titanium forging, there may not be enough material within one forging, and forg-

ing may interact with experimental variables. The operator's role in this highly automated process is small, and material properties of the blank titanium forgings are carefully controlled because of the criticality of the part.

6. NUISANCE FACTORS

Processes vary over time. Experimental conditions vary over time. "Identical samples" differ. Some variations are innocuous, some are pernicious. Examples include contamination of process fluids over time, equipment wear, build-up of oxides on tools, and so forth. Some of these can be measured and monitored to at least ensure that they are within limits; others must be assessed subjectively by experts; still others are unmeasured. Nuisance factors are not controlled, and are not of primary interest in this experiment. They differ from held-constant factors in that they cannot be deliberately set to a constant level for all experimental units. If the level can be selected for any experimental unit, however, blocking or randomization might be appropriate. If levels cannot be selected (i.e., the levels of the factor are unpredictable, perhaps continuous), then the nuisance factor becomes a covariate in the analysis. If a nuisance factor is not measurable and thought to be very influential, it may also be called an *experimental risk* factor. Such factors can inflate experimental error, making it more difficult to assess the significance of control variables. They can also bias the results. For the CNC-machining example, the nuisance factors are as shown in Table 5.

Experiment designers have a set of passive strategies (randomization, blocking, analysis of covariance, stratified analysis) to reduce the impact of nuisance factors. These strategies can have a major effect on the experimental design. They may be constrained

Table 4. Held-Constant Factors

Factor (units)	Desired experimental level and allowable range	Measurement precision—how known?	How to control (in experiment)	Anticipated effects
Type of cutting fluid	Standard type	Not sure, but thought to be adequate	Use one type	None
Temperature of cutting fluid (degrees F.)	100–110°F. when machine is warmed up	1–2° F. (estimate)	Do runs after machine has reached 100°	None
Operator	Several operators normally work in the process	—	Use one "mid-level" operator	None
Titanium forgings	Material properties may vary from unit to unit	Precision of lab tests unknown	Use one lot (or block on forging lot, only if necessary)	Slight

Table 5. Nuisance Factors

Nuisance factor (units)	Measurement precision—how known?	Strategy (e.g., randomization, blocking, etc.)	Anticipated effects
Viscosity of cutting fluid	Standard viscosity	Measure viscosity at start and end	None to slight
Ambient temperature (°F.)	1–2° F. by room thermometer (estimate)	Make runs below 80°F.	Slight, unless very hot weather
Spindle	—	Block or randomize on machine spindle	Spindle-to-spindle variation could be large
Vibration of machine during operation	?	Do not move heavy objects in CNC machine shop	Severe vibration can introduce variation within an impeller

by limits on the number of observations, costs of changing control-variable settings, and logistic considerations. In the CNC-machining example, the only nuisance factor to have potentially serious effects and for which blocking seems appropriate is the machine spindle effect (though it may be necessary to also block on titanium forgings). The machine has four spindles, requiring a design with four blocks or randomizing on all four. Blocking will introduce a bias in the estimates confounded with the blocking variable(s), whereas randomization will inflate the experimental error. The other two factors are dealt with by ensuring that they stay below levels at which problems may be encountered.

7. INTERACTIONS

The interactions sheet is self-explanatory. Unfortunately, the concept of interactions is *not* self-explanatory—even among intelligent, mathematically inclined people in the sciences. Hence, as part of the package of guide sheets, it is helpful to include a tutorial. The graphical portion of the tutorial is presented in Figure 6. An additional, expository description of interactions is sometimes included, but it is not shown here. The interactions table explicitly recognizes only pairwise interactions of linear terms. It provides an opportunity for the experimenters to capture knowledge or speculation that certain pairwise interactions may be present and others are unlikely to be present. This input is helpful when the experiment is later designed—to choose resolution, or more generally to choose which effects should or should not be confounded. Higher order effects may also be important but are not captured in the guide sheets. The interaction sheet for the CNC-machining example is shown in Table 6.

A helpful way to use this matrix is to *avoid* discussing every possible pairwise interaction one at a time but instead use the process of elimination or inclusion; that is, if interactions are generally im-

portant, a question can be posed: "Are there any interactions that are arguably *not* present?" If main effects dominate interactions, a question can be posed: "Are there any interactions that must be estimated clear of main effects?" Alternatively, a secret-ballot vote on potentially important interactions can be held among experimenters and other knowledgeable investigators, with each receiving, say, 100 votes to spread among the interactions.

The remaining items are found on the Master Guide Sheet.

8. RESTRICTIONS, PREFERENCES FOR THE DESIGN, ANALYSIS, AND PRESENTATION

Box, Hunter, and others have repeatedly exhorted, "Attention to detail can determine the success or failure of the experiment." Item 8 in Figure 2 is part of heeding that advice. *Theoretical* optimal experimental design and *practical* experimental design are often worlds apart, and restrictions often make the difference. Since a single unknown restriction can render worthless an otherwise well-considered, laboriously developed design, the statistician should encourage experimenters to be quick to put these limitations and pitfalls on the table. In particular, there appears to be a lack of awareness in the applied statistics community of the prevalence of experiments with unidentified split-plot structure. Because it is unidentified (different experimental units used for different parts of the experiment), the analysis is often done incorrectly—using the wrong error terms to test statistical significance. The discussion of the issues surrounding the choice of experimental unit and analysis strategies goes beyond the scope of this article but should take place on the experimentation team.

Items 10 and 11 of Figure 2 are intended for the following three circumstances: First, when experimenters are statistically sophisticated and have a good idea of appropriate designs or analysis techniques;
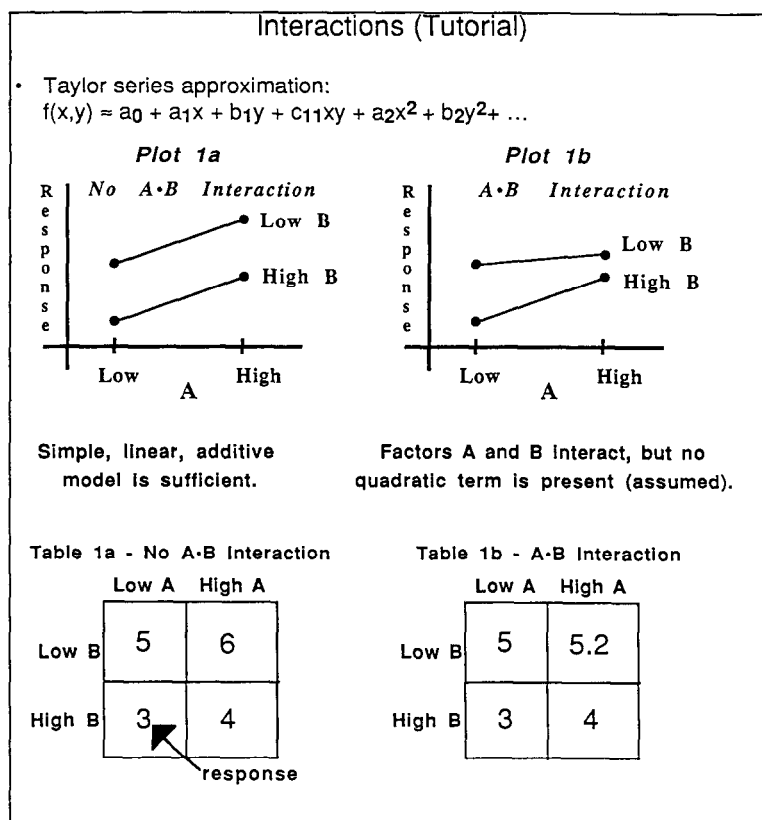
Figure 6. Graphical (tutorial) Presentation of Interactions.

second, when the experiment has been preceded by experiments in which a particular design or technique proved to be useful; third, when, on considering designs, analyses, and plots, the experimenters may want to change information in items 2–7—for example, narrowing the scope of the objective or increasing the number of settings for a control variable.

9. THE NEXT STAGE

By the time the experimentation team has come to a consensus concerning the information collected in items 1–10 of the guide sheet, the statistician (or surrogate) will have had the opportunity to step beyond the generic confines of the guide sheet and discuss more problem-specific issues that will affect the experimental design, such as multilevel factors,

different sizes of experimental units, and logistics. Then, it may be useful to (a) choose candidate designs, (b) review them with the experimenters in the context of the collected information to determine if any of the designs should be dropped from further consideration, and (c) write an experimental design proposal that contains (at least) one or more proposed designs; a comparative analysis of the designs with respect to number of runs, resolution (or aliased effects), number of distinct control variable combinations, prediction error standard deviation, and so forth; a design recommendation with justification; and copies of the completed guide sheets.

When an experimental design has been selected, the sheets are used to help launch supportive tasks required for the experiment to be successful. This

Table 6. Interactions

Control variable	y shift	z shift	Vendor	a shift	Speed	Height	Feed
x shift			P				
y shift	—		P				
z shift	—	—	P				
Vendor	—	—	—	P			
a shift	—	—	—	—			
Speed	—	—	—	—	—		F, D
Height	—	—	—	—	—	—	

NOTE. Response variables are P = profile difference, F = surface finish, and D = surface defects.

involves issues addressed in items 11 and 12 on the guide sheet. Additionally, there will be logistical planning and planning for measurement capability studies, process capability studies, preexperiments to quantify the effects of various factors (held-constant and nuisance) on response variables, and trial runs.

In regard to item 12 of Figure 2, an experiment without a coordinator will probably fail. Though a statistician can play this role, it is often better played by another member of the experimentation team, who can "champion" the experiment among peers. The statistician (or surrogate) can play a strong support role and be primarily responsible for that in which he or she is professionally trained—the design and analysis of the experiment and *not* the execution.

Finally, considering item 13 of the guide sheet, the team should entertain the idea of trial runs to precede the experiment—especially if this is the first in a series of experiments. A trial run can consist of a centerpoint run or a small part (perhaps a block) of the experiment. The first and most important purpose of trial runs is to learn and refine experimental procedures without risking the loss of time and expensive experimental samples. Most experiments involve people (and sometimes machines) doing things that they have *never done before*. Usually some practice helps.

A second important reason for trial runs is to estimate experimental error before expending major resources. An unanticipated large experimental error could lead to canceling or redesigning the experiment, widening the ranges of settings, increasing the number of replicates, or refining the experimental procedure. An unanticipated small experimental error (does this ever really happen?) could have opposite effects on plans or cause the experimenters to reassess whether the estimate is right or complete.

A third reason is that trial runs are also excellent opportunities to ensure that data-acquisition systems are functioning and will permit experimental runs to be conducted as fast as had been planned.

Last, a fourth reason is that trial runs may yield results so unexpected that the experimenters decide to change their experimental plans.

Naturally, the feasibility and advisability of conducting trial runs depends on the context, but the experiment teams in which we have been involved have *never regretted* conducting trial runs. Some trial runs have saved experiments from disaster.

10. SUMMARY

To conduct complex experiments, careful planning with attention to detail is critical. Predesign planning is one part of the process by which experiments are conceived, planned, executed, and interpreted. It is often the part claimed by no one, hence it is often done informally—and sloppily. The use of predesign experiment guide sheets provides a way to systematize the process by which an experimentation team does this planning, to help people to (a) more clearly define the objectives and scope of an experiment and (b) gather information needed to design an experiment.

ACKNOWLEDGMENTS

REFERENCES

Bishop, T., Petersen, B., and Trayser, D. (1982), "Another Look at the Statistician's Role in Experimental Planning and Design," *The American Statistician*, 36, 387–389.

Box, G. E. P., Hunter, W. G., and Hunter, J. S. (1978), *Statistics for Experimenters*, New York: John Wiley.

Eisenhart, C. (1962), "Realistic Evaluation of the Precision and Accuracy of Instrument Calibration Systems," *Journal of Research of the National Bureau of Standards*, 67C, 161–187.

Hahn, G. (1977), "Some Thing Engineers Should Know About Experimental Design," *Journal of Quality Technology*, 9, 13–20.

———— (1984), "Experimental Design in a Complex World," *Technometrics*, 26, 19–31.

Hoadley, A., and Kettenring, J. (1990), "Communications Between Statisticians and Engineers/Physical Scientists" (with commentary), *Technometrics*, 32, 243–274.

Hunter, W. G. (1977), "Some Ideas About Teaching Design of Experiments With 2^5 Examples of Experiments Conducted by Students," *The American Statistician*, 31, 12–17.

McCulloch, C. E., Boroto, D. R., Meeter, D., Polland, R., and Zahn, D. A. (1985), "An Expanded Approach to Educating Statistical Consultants," *The American Statistician*, 39, 159–167.

Montgomery, D. C. (1991), *Design and Analysis of Experiments* (3rd ed.) New York, John Wiley.

Natrella, M. G. (1979), "Design and Analysis of Experiments," in *Quality Control Handbook*, ed. J. M. Juran, New York: McGraw-Hill, pp. 27-35.