

Evaluación de la Robustez del sistema Mahalanobis-Taguchi a diferentes Arreglos Factoriales

Jorge Limón⁽¹⁾, Manuel A. Rodríguez⁽²⁾, Yolanda A. Báez⁽¹⁾ y Diego A. Tlapa⁽¹⁾

(1) Universidad Autónoma de Baja California, Facultad de Ingeniería Ensenada, Km. 103 Carretera Tijuana-Ensenada, Baja California-México (e-mail: jorge.limon@uabc.edu.mx)

(2) Instituto Tecnológico de Ciudad Juárez, Blvd. Tecnológico No. 1340, Ciudad Juárez, Chihuahua-México.

Recibido Dic. 03, 2010; Aceptado Ene. 20, 2011; Versión Final recibida Ene. 30, 2011

Resumen

Se evalúa la robustez del Sistema Mahalanobis-Taguchi a los diferentes arreglos que pudieran utilizarse para discriminar las variables consideradas en un estudio. Para esto se utilizaron diferentes niveles de fraccionado de un diseño factorial 2^9 , así como todas las fracciones posibles para cada nivel, para evaluar si el resultado variaba dependiendo del arreglo empleado. Se utilizaron para este análisis los datos del estudio del cáncer de mama de Wisconsin reportados en una publicación, en cuyo análisis utilizaron un arreglo ortogonal L_{12} . En este trabajo se realizó el análisis con diseños 2^{9-k} ($k = 0, 1, 2, 3, 4$ y 5) y todas las fracciones posibles para cada valor de k , generadas con el software Minitab 15 y se generó un programa en Matlab. Los resultados demostraron que esta técnica no es robusta a los diferentes arreglos que pudieran utilizarse.

Palabras clave: robustez, sistema Mahalanobis-Taguchi, diseño factorial, arreglo ortogonal

A robustness evaluation of the Mahalanobis-Taguchi System to different Factorial Arrays

The robustness of the Mahalanobis-Taguchi System to different arrays that could be used to discriminate variables in a study, is evaluated.. For this, different levels of fractional factorial design 2^9 were used, as well as all possible fractions for each level to find out if the results varied depending on the array utilized. For this analysis the Wisconsin's breast cancer data reported in a publication were used, in which, an orthogonal array L_{12} was employed. In this work, the analysis was performed with 2^{9-k} designs ($k = 0, 1, 2, 3, 4$ and 5) and all the possible fractions for each value of k generated using the software Minitab 15 and a Matlab program was generated. The results showed that this method is not robust to the different arrays that could be used.

Keywords: robustness, Mahalanobis-Taguchi system, factorial design, orthogonal array

INTRODUCCIÓN

Las técnicas y métodos utilizados para la optimización y análisis de variables han tenido un comportamiento creciente durante las últimas décadas con la aplicación y adaptación de principios estadísticos, enfocados al análisis de datos con comportamiento multivariante. En la actualidad se han desarrollado diferentes tecnologías de reconocimiento de patrones para la realización de análisis y predicciones precisas en sistemas multidimensionales. Taguchi y Jugulum (2002) indicaron entre otras las siguientes: análisis de componentes principales (ACP), Método de discriminación y clasificación, regresión múltiple, redes neuronales artificiales (RNA) y el sistema Mahalanobis-Taguchi (SMT). Otros autores como Jain et al. (2000) mencionaron otros enfoques como análisis discriminante lineal y árboles de decisión de regresión logística.

En el presente trabajo se analiza el SMT, el cual ha sido utilizado en diferentes aplicaciones de diagnóstico para generar indicadores cuantitativos para la toma de decisiones por medio de la construcción de una escala de medición multivariable. Según Cudney et al. (2009), este sistema emplea la distancia Mahalanobis (DM), la cual es una medida basada en las correlaciones entre variables y diferentes patrones que pueden ser identificadas y analizadas con respecto a un grupo base o de referencia. Estas correlaciones son una medición aproximada de la interacción entre las variables, es por eso que la DM es usada para representar patrones correspondientes a combinaciones de las variables (Taguchi et al., 2005). Taguchi llevó al desarrollo de SMT, al proveer medios para definir el grupo de referencia y medir el grado de anormalidad de observaciones individuales (Taguchi y Jugulum, 2000). El SMT construye una escala de medición multivariable para entender el comportamiento de diferentes patrones, lo que a su vez ayudará a medir o predecir varias condiciones del sistema multivariable, para que el analista pueda tomar acciones correctivas apropiadas (Taguchi et al., 2005).

Jugulum y Samuel (2008), reportaron que existe una considerable cantidad de literatura sobre SMT con más de 500 aplicaciones. Sin embargo, debido precisamente a esta gran cantidad de aplicaciones y al creciente interés por esta herramienta, es que para este trabajo se ha considerado necesario hacer una evaluación para conocer su comportamiento en diferentes escenarios que pueden presentarse, a la hora de hacer el análisis de datos mediante la selección de diferentes arreglos factoriales para alojar las variables consideradas en un estudio para determinar su importancia. Esto para darle información a considerar a los posibles usuarios en el momento de escoger la herramienta apropiada para el tratamiento de los datos. Los resultados encontrados en este trabajo mostraron que el SMT es poco confiable bajo las circunstancias antes mencionadas.

Los métodos Taguchi (MT), han resultado de alto impacto en ingeniería de calidad debido al mejoramiento significativo de las características de calidad relacionadas al desarrollo de productos y procesos (Taguchi, 1991). Dentro de su filosofía, Taguchi muestra como el diseño estadístico de experimentos puede ayudar a los ingenieros a diseñar y manufacturar productos que sean de alta calidad y a un bajo costo. Su enfoque se basa primeramente en eliminar las causas de pobre calidad y en hacer que el funcionamiento de los productos sea insensible a la variación (Antony y Antony, 2001). En los MT, los diseños factoriales fraccionados llamados arreglos ortogonales, son utilizados para optimizar la cantidad de información obtenida y limitar el número de experimentos. Los AOs son descritos por columnas que representan los factores a estudiar y los renglones representan los experimentos individuales. En la literatura se pueden encontrar muchas aplicaciones exitosas de los MT, por ejemplo en Báez et al. (2010).

Al trabajar con sistemas multidimensionales, el total de combinaciones a examinar puede llegar a ser muy grande, por lo que resulta difícil de examinarlas todas, en su lugar, Taguchi y Jugulum (2002), propusieron el uso de AOs para reducir la cantidad de combinaciones a ser probadas en el SMT, donde las variables significativas o útiles son identificadas para futuros análisis o diagnósticos de reconocimiento de patrones. Los AOs y la razón señal a ruido (S/R) son utilizados para identificar estas variables significativas. Al respecto Woodall et al. (2003) mencionaron que los AOs son usados para encontrar las combinaciones de variables que maximizan la razón S/R, por otro lado, comentan que SMT, resulta en una escala de medida relacionada con la distancia de Mahalanobis para definir la anormalidad presentada entre las variables anormales y las

variables clasificadas como normales, demostrando por principio que la DM basada en todas las variables disponibles, es capaz de separar las variables anormales de las normales. La optimalidad es definida en términos de la capacidad de la escala de la DM de emparejar una escala pre especificada para medir el impacto de las variables anormales.

En una aplicación del SMT, Cudney et al. (2007) emplearon un AO L_8 para minimizar el número de corridas. En ese estudio, seis parámetros de calidad y rendimiento del manejo de un vehículo fueron evaluados, encontrando que sólo cinco fueron útiles. Adicionalmente Cudney et al. (2008), presentaron otro trabajo utilizando SMT acerca del índice de satisfacción de cliente y el funcionamiento de vehículos (factores de sensibilidad del pedal de freno) para determinar las variables útiles. Para realizar el análisis, los autores emplearon un L_{16} para nueve índices de sensibilidad del pedal de freno. Por otra parte, en un aspecto similar al presente trabajo, Bovas y Asokan (2003), encontraron que las variables significativas encontradas con SMT son diferentes cuando la ubicación de las variables en las columnas de los arreglos ortogonales también es diferente y mencionan que esta inconsistencia en la identificación de las variables útiles es debido al hecho de que la interacción entre variables no es tomada en cuenta, y también debido a que la selección de variables útiles es basada simplemente en efectos principales positivos (ganancia) y no en alguna prueba estadística formal.

En el presente estudio, se analizó el trabajo de Cudney et al. (2007), en el cual se utilizó un AO L_{12} y llevó a eliminar tres atributos que no fueron significativos en la exactitud de diagnóstico (espesor de tumor, adhesión marginal y tamaño de células epiteliales). Dicho trabajo se basó en los datos del cáncer de seno del UCI machine-learning repository, los cuales fueron recolectados en la universidad de Wisconsin por W.H. Wolberg (1991). Para el desarrollo de este trabajo, se consideró la posibilidad de que para realizar un estudio, diferentes investigadores pudieran contar con diferentes recursos para hacer el análisis del problema, lo que pudiera afectar la selección del arreglo a utilizar y con ello la cantidad de corridas (combinaciones de presencia – ausencia de los factores), de tal forma que si quizá los cálculos han de hacerse manualmente, tal vez se opte por el diseño más pequeño que comprenda todos los factores considerados. Sin embargo, si la cantidad de recursos disponibles para realizar los cálculos aumenta, de estar trabajando con un diseño 2^{k-p} se pudiera optar por trabajar con una fracción un poco más grande, por ejemplo la fracción 2^{k-p+1} , y así estar aumentando la fracción de acuerdo a los recursos de análisis disponibles, hasta llegar quizá a utilizar el arreglo factorial 2^k completo. Taguchi, et al. (2005) comentan que en sistemas multidimensionales el número total de combinaciones a analizar puede ser del orden de varios miles, por lo que según mencionan, no es posible examinarlas todas, por lo que recomiendan el uso de AOs para reducir el número de combinaciones a ser probadas. A decir verdad cuando los cálculos se hacen utilizando programas especializados de computadora, esto último deja de ser relevante, razón por la cual en este trabajo se tuvo la posibilidad de analizar todas las fracciones posibles e incluso el arreglo 2^9 completo.

METODOLOGÍA

En el SMT, el espacio de Mahalanobis (EM) se obtiene utilizando las variables estandarizadas de los datos sanos o normales, y se utiliza para discriminar entre los objetos normales y anormales. Una vez que se establece el EM, el número de atributos se reduce utilizando los AOs y la razón señal a ruido, evaluando la contribución de cada uno de ellos. Cada fila del AO determina un subconjunto del sistema original incluyendo y excluyendo ese atributo del sistema. Las diferentes etapas del método SMT se resumen a continuación:

Etapas I: Construcción de una escala de medición: i) Seleccionar un grupo de referencia con variables y observaciones tan uniformes como sea posible: ii) Utilizar a este grupo de referencia como una base o punto de referencia de la escala.

Etapas II: Validación de la escala de medición: i) Identificar las condiciones fuera del grupo de referencia: ii) Calcular las DMs de estas condiciones y comprobar si coinciden con el juicio de quien tomará las decisiones: iii) Calcular las razones señal a ruido para determinar la exactitud de la escala.

Etapa III: Identificación de las variables útiles (etapa de desarrollo): i) Determinar el grupo de variables útiles utilizando AOs y la razón señal a ruido.

Etapa IV: Diagnóstico futuro con las variables útiles: i) Monitorear las condiciones utilizando la escala, la cual se desarrolla con la ayuda del grupo de variables útiles; ii) Basados en los valores de las DMs, se pueden tomar las acciones correctivas apropiadas.

Sistema Mahalanobis-Taguchi (SMT)

El primer paso en MTS es construir una escala de medición utilizando la DM como referencia. Para construir una escala de medición, es necesario recoger un conjunto de observaciones normales, las cuales se estandarizan utilizando la ecuación (1).

$$Z_i = \frac{X_i - m}{\sigma} \quad (1)$$

Donde m representa la media del atributo, σ es la desviación estándar del atributo, Z_i representa las variables estandarizadas y X_i representa las observaciones normales. El vector estandarizado se obtiene de los valores estandarizados de x_i ($i=1,2,3,\dots,k$). La DM mide la distancia en espacios multidimensionales tomando en cuenta la correlación entre los atributos. El significado estadístico de la DM es la proximidad de un punto desconocido a la media del grupo. La ecuación (2) se utiliza para calcular las DMs:

$$MD_j = D_j^2 = \frac{1}{k} Z_{ij}^T C^{-1} Z_{ij} \quad (2)$$

Donde C^{-1} es la inversa de la matriz de correlación, la cual contiene los coeficientes de correlación entre variables, T es la transpuesta del vector estándar y k es el número de variables. Es fácil comprobar que el valor medio de la DMs es 1 para todas las observaciones en el EM. Por esta razón, el SM es llamado espacio unitario (Taguchi y Jugulum, 2002). Sin embargo, estudios recientes mencionan que utilizando algebra matricial se puede demostrar que el valor medio de la DM para los g datos del grupo normal es siempre exactamente $(g - 1) / g$ (Woodall y Koudelik, 2003).

El segundo paso es validar la escala de medición. Para esto, se utilizan las observaciones fuera de la DM, generalmente observaciones anormales o de prueba. Para calcular la DM de las observaciones anormales se utiliza el valor medio, la desviación estándar y la matriz de correlación de las observaciones normales. En las escalas de medición correctas, las DMs de las observaciones anormales son más grandes que las DMs de las observaciones normales. El tercer paso del SMT es optimizar el sistema. Para este fin, son de gran utilidad los AOs y la razón S/R para identificar los atributos importantes. En el experimento, cada factor se asigna a una columna en el AO, y cada fila representa la combinación experimental de una corrida. Se utiliza un AO de dos niveles para representar presencia o ausencia, donde el nivel 1 corresponde a la presencia de la variable y el nivel 2 indica la ausencia de la variable. Cada atributo será utilizado o descartado con respecto al AO y la razón S/R es calculada.

Existen diferentes tipos de razón S/R; sin embargo, el SMT utiliza lo mayor es lo mejor o la razón señal a ruido dinámica. En el contexto del SMT, la razón señal a ruido se define como la medida de exactitud de la predicción de la escala. Esta refleja la severidad de las anomalías y la diferencia de los promedios en los valores de la razón señal a ruido de cada atributo cuando es incluido o excluido. Para un atributo dado X_i , SR^+ representa la razón señal a ruido promedio de incluir el atributo X_i . Por otra parte SR^- representa cuando x_i es excluido.

$$Ganancia = SR^+ - SR^- \quad (3)$$

Si el aumento en la ecuación (3) es positivo, se utiliza el atributo, si no, éste es excluido. Se realiza una corrida de confirmación construyendo un EM utilizando las variables útiles. Las DMs de las observaciones anormales también se calculan tomando como base las variables útiles. La DM promedio del grupo normal es comparada con el promedio de la DM del grupo anormal.

Caso de estudio

Este estudio es referido a Cudney et al., (2007), donde su meta es predecir si una muestra de tejido, recogida del pecho de pacientes es maligna o benigna. Para propósitos de nuestro trabajo se utilizará la misma muestra de tamaño 30, de la cual 19 resultados fueron benignos y 11 malignos, considerando los mismos 9 atributos a evaluar. En esta parte de la investigación se muestran los resultados de las distancias de Mahalanobis obtenidas utilizando un AO L_{12} como originalmente fueron hechos en el trabajo de referencia. Los datos, así como los resultados de las DMs pueden verse en la tabla 3.

En las tablas 1 y 2 se muestran la matriz de correlaciones y la matriz de correlaciones inversa respectivamente, utilizadas para calcular las DMs. En la tabla 4 se muestra el AO L_{12} y las razones S/R obtenidas para combinación de factores del arreglo y en la tabla 5 se muestran las razones S/R por nivel, así como las ganancias correspondientes a cada uno de los 9 atributos considerados. Después de esto, los mismos cálculos para la fracción 9 (F9) de las 16 posibles fracciones del arreglo factorial 2^{9-4} se hacen para realizar una comparación de los resultados. Esta fracción se formó con los generadores F=-BCDE, G=-ACDE, H=-ABDE y J=ABCE.

Tabla 1: Matriz de correlaciones

Matriz de correlaciones								
1.00	-0.27	0.27	0.25	0.32	-0.13	-0.15	0.07	0.18
-0.27	1.00	-0.07	-0.16	-0.17	-0.09	0.05	0.29	0.36
0.27	-0.07	1.00	0.21	0.02	0.02	-0.13	-0.29	0.22
0.25	-0.16	0.21	1.00	0.79	0.47	-0.14	-0.15	-0.02
0.32	-0.17	0.02	0.79	1.00	0.36	0.16	0.19	-0.10
-0.13	-0.09	0.02	0.47	0.36	1.00	-0.11	-0.22	-0.10
-0.15	0.05	-0.13	-0.14	0.16	-0.11	1.00	0.23	-0.27
0.07	0.27	-0.29	-0.15	0.19	-0.22	0.23	1.00	-0.12
0.18	0.36	0.22	-0.02	-0.10	-0.10	-0.27	-0.12	1.00

Tabla 2: Matriz de correlaciones inversa

Matriz de correlaciones inversa								
1.62	0.54	-0.37	0.10	-0.68	0.40	0.28	-0.25	-0.38
0.54	1.68	-0.08	-0.46	0.59	-0.05	-0.21	-0.76	-0.80
-0.37	-0.01	1.30	-0.56	0.48	0.03	-0.15	0.24	-0.19
0.10	-0.46	-0.56	4.64	-3.98	-0.35	1.04	1.16	0.34
-0.68	0.59	0.48	-3.98	5.06	-0.49	-1.25	-1.42	-0.32
0.40	-0.05	0.03	-0.35	-0.49	1.50	0.21	0.33	0.13
0.28	-0.21	-0.15	1.04	-1.25	0.21	1.47	0.15	0.39
-0.25	-0.76	0.24	1.16	-1.42	0.33	0.15	1.85	0.44
-0.38	-0.80	-0.19	0.34	-0.32	0.13	0.39	0.44	1.54

Como se puede apreciar en la tabla 3, en el caso del comportamiento de los datos normales (saludables), el valor de la DM es muy pequeño y el promedio de la DM está cercano a uno. Los valores anormales de la DM son más grandes que los normales, lo que ilustra la capacidad de clasificación de la DM. Finalmente usando un AO L_{12} y la razón señal a ruido lo mayor es lo mejor, se prueba la importancia de cada atributo. Los datos de prueba anormales se utilizan para calcular la razón señal a ruido.

Tabla 3: Datos seleccionados del cáncer de mama y DM (OA L₁₂)

Corrida	Atributo										Clase	DM
	A	B	C	D	E	F	G	H	I			
1	5	1	1	3	4	1	3	2	1	2	1.75	
2	5	1	2	10	4	5	2	1	1	2	1.61	
3	1	1	1	1	2	1	2	1	1	2	0.57	
4	1	1	1	1	2	5	1	1	1	2	1.64	
5	5	1	1	6	3	1	1	1	1	2	0.85	
6	5	1	1	1	2	1	2	2	1	2	0.84	
7	3	1	1	1	2	1	3	1	1	2	0.40	
8	4	1	2	1	2	1	3	1	1	2	0.44	
9	5	1	1	1	2	1	1	1	1	2	0.47	
10	5	1	1	1	2	2	2	1	1	2	0.39	
11	4	1	3	3	2	1	1	1	1	2	0.94	
12	5	2	2	2	2	1	1	1	2	2	1.90	
13	3	1	1	3	2	1	1	1	1	2	0.44	
14	5	1	3	1	2	1	2	1	1	2	0.89	
15	5	1	1	1	2	1	2	2	1	2	0.84	
16	1	1	1	2	2	1	3	1	1	2	0.86	
17	1	3	1	1	2	1	2	2	1	2	1.68	
18	4	2	1	1	2	2	3	1	1	2	1.04	
19	5	1	1	1	2	1	1	1	1	2	0.47	
20	10	10	8	10	6	5	10	3	1	4	75.80	
21	10	10	10	7	9	10	7	10	10	4	333.00	
22	7	9	4	10	10	3	5	3	3	4	71.82	
23	5	10	10	8	5	5	7	10	1	4	119.50	
24	5	5	5	2	5	10	4	3	1	4	28.76	
25	8	6	5	4	3	10	6	1	1	4	33.50	
26	8	4	4	1	2	9	3	3	1	4	19.76	
27	4	2	3	5	3	8	7	6	1	4	43.02	
28	6	1	3	1	4	5	5	10	1	4	85.88	
29	10	4	7	2	2	8	6	1	1	4	28.15	
30	9	5	8	1	2	3	2	1	5	4	43.34	

Tabla 4: Arreglo Ortogonal L₁₂ y Razón S/R

Corrida	Atributos										Razón S/R
	A	B	C	D	E	F	G	H	I		
1	1	1	1	1	1	1	1	1	1	1	31.60
2	1	1	1	1	1	2	2	2	2	2	16.45
3	1	1	2	2	2	1	1	1	2	2	31.12
4	1	2	1	2	2	1	2	2	1	1	20.11
5	1	2	2	1	2	2	1	2	1	2	13.90
6	1	2	2	2	1	2	2	1	2	1	16.07
7	2	1	2	2	1	1	2	2	1	2	18.27
8	2	1	2	1	2	2	2	1	1	1	26.05
9	2	1	1	2	2	2	1	2	2	1	26.18
10	2	2	2	1	1	1	1	2	2	1	07.24
11	2	2	1	2	1	2	1	1	1	2	26.52
12	2	2	1	1	2	1	2	1	2	2	29.79

En la tabla 4 se muestran el AO L_{12} y las razones señal a ruido. La tabla 5 muestra los resultados de la optimización. Los atributos que tienen ganancia negativa se pueden considerar como no útiles y sacarse del estudio. Por otra parte los factores con ganancia positiva deben considerarse para análisis posteriores, siendo estos los atributos B, C, F, G, H e I. Ahora se presentan los mismos cálculos hechos anteriormente, pero utilizando la fracción 9 (F9) de las 16 posibles fracciones del arreglo factorial 2^{9-4} , la cual se formó con los generadores mencionados anteriormente. Cuando se calcularon las DMs para los datos normales y anormales estas coincidieron con las obtenidas antes, lo cual es obvio porque se usaron los mismos datos y aún no se ha utilizado el arreglo factorial, así que estos resultados se omiten porque son iguales a los reportados en la tabla 3. Una vez que se utilizó el arreglo factorial 2^{9-4} F9 en la fase de optimización, se obtuvieron las razones señal a ruido, las cuales se muestran en la tabla 6.

Tabla 5: Razones señal a ruido por nivel y ganancias

	A	B	C	D	E	F	G	H	I
Nivel 1	129.30	149.70	150.70	125.00	116.20	138.10	136.60	161.20	136.50
Nivel 2	134.10	113.60	112.70	138.30	147.20	125.20	126.70	102.20	126.90
Ganancia	-4.80	36.04	38.00	-13.20	-31.00	12.96	9.82	59.00	9.60

Tabla 6: Diseño Factorial Fraccionado 2^{9-4}_{VI} F9 y Razón S/R

Corrida	Atributos									Razón S/R
	A	B	C	D	E	F	G	H	I	
1	1	1	1	1	1	1	1	1	2	31.29
2	2	1	1	1	1	1	2	2	1	24.40
3	1	2	1	1	1	2	1	2	1	20.37
4	2	2	1	1	1	2	2	1	2	25.93
5	1	1	2	1	1	2	2	1	1	27.86
6	2	1	2	1	1	2	1	2	2	22.72
7	1	2	2	1	1	1	2	2	2	19.69
8	2	2	2	1	1	1	1	1	1	27.51
9	1	1	1	2	1	2	2	2	2	16.38
10	2	1	1	2	1	2	1	1	1	29.15
11	1	2	1	2	1	1	2	1	1	29.48
12	2	2	1	2	1	1	1	2	2	26.20
13	1	1	2	2	1	1	1	2	1	23.82
14	2	1	2	2	1	1	2	1	2	29.35
15	1	2	2	2	1	2	1	1	2	19.20
16	2	2	2	2	1	2	2	2	1	-11.34
17	1	1	1	1	2	2	2	2	1	13.31
18	2	1	1	1	2	2	1	1	2	29.28
19	1	2	1	1	2	1	2	1	2	28.16
20	2	2	1	1	2	1	1	2	1	25.44
21	1	1	2	1	2	1	1	2	2	25.80
22	2	1	2	1	2	1	2	1	1	29.51
23	1	2	2	1	2	2	1	1	1	19.67
24	2	2	2	1	2	2	2	2	2	-39.07
25	1	1	1	2	2	1	1	1	1	31.72
26	2	1	1	2	2	1	2	2	2	22.39
27	1	2	1	2	2	2	1	2	2	23.03
28	2	2	1	2	2	2	2	1	1	27.74
29	1	1	2	2	2	2	2	1	2	30.14
30	2	1	2	2	2	2	1	2	1	23.08
31	1	2	2	2	2	1	2	2	1	16.68
32	2	2	2	2	2	1	1	1	2	04.51

Con la información reportada en la tabla anterior, los efectos (ganancias) para las variables son calculados y se muestran en la tabla 7. Estos efectos sugieren que todos los factores son importantes (ganancias positivas), excepto el factor D (ganancia negativa), por lo que el resultado es diferente al obtenido con el AO L_{12} . Posteriormente se muestra el resumen de los resultados obtenidos en cuanto a las ganancias para todos los arreglos factoriales utilizados en la fase de optimización en este trabajo, para evaluar sí el resultado en cuanto a la determinación de los factores importantes se mantenía o variaba, dependiendo del arreglo empleado.

Tabla 7: Razones señal a ruido por nivel y ganancias (F9)

	A	B	C	D	E	F	G	H	I
Nivel 1	376.63	410.19	404.28	331.89	362.03	395.96	382.80	420.51	358.40
Nivel 2	296.79	263.23	269.14	341.52	311.38	277.46	290.62	252.91	315.01
Ganancia	79.84	146.96	135.14	-9.63	50.65	118.50	92.18	167.60	43.39

Tabla 8: Ganancias utilizando las 32 fracciones posibles del arreglo 2^{9-5}

F 2^{9-5}	A	B	C	D	E	F	G	H	I
F1	9.85	44.68	36.59	-41.68	-23.02	51.96	50.11	91.28	9.19
F2	49.33	70.91	61.72	-64.86	50.08	77.58	78.03	113.62	47.44
F3	-26.46	73.28	58.85	27.92	22.54	67.13	68.36	106.07	39.77
F4	7.72	32.54	27.73	3.18	3.75	43.95	32.48	74.96	11.06
F5	41.95	37.36	78.06	-8.96	-0.99	88.38	14.95	107.19	29.72
F6	12.87	57.66	56.53	-27.43	20.13	63.38	-0.98	76.10	13.30
F7	40.54	65.97	65.51	-1.30	-47.92	49.92	37.01	55.94	-19.03
F8	34.17	60.43	72.64	-0.48	46.74	59.48	47.25	51.42	-16.68
F9	9.19	47.56	76.19	-26.01	-27.44	61.62	52.17	24.37	16.62
F10	4.72	42.59	64.40	-16.89	33.82	70.70	58.20	28.96	30.38
F11	41.95	29.64	81.77	-28.43	-10.69	48.74	57.02	67.34	-9.75
F12	17.03	47.26	59.86	-9.12	18.02	32.63	35.29	49.41	15.33
F13	-5.72	66.21	67.56	-42.33	-11.58	61.69	3.62	38.29	9.53
F14	9.89	59.40	55.59	-35.94	13.07	59.22	6.87	43.51	21.15
F15	30.62	70.58	71.08	10.31	-35.29	43.80	20.31	60.98	-7.58
F16	51.40	81.88	76.29	25.40	48.65	54.77	37.40	70.70	-12.52
F17	34.81	67.17	65.20	-18.67	-37.41	34.71	24.90	61.92	5.88
F18	56.34	82.69	83.49	-27.78	46.69	45.01	34.15	66.79	-13.34
F19	-12.23	87.05	80.85	52.95	3.08	88.16	15.11	48.62	3.92
F20	16.89	67.84	67.43	30.18	17.42	68.56	-5.66	33.75	16.12
F21	55.74	26.41	86.33	26.86	1.69	60.18	56.53	68.84	-4.56
F22	25.31	52.71	70.98	3.79	20.20	35.82	34.02	39.41	10.25
F23	17.94	50.91	67.05	18.23	-27.09	61.77	59.23	32.02	28.47
F24	11.06	46.28	69.71	9.11	26.59	64.18	66.95	32.54	22.88
F25	36.72	64.18	72.74	-8.80	-40.37	53.92	32.53	49.04	-8.86
F26	31.85	60.34	72.33	-1.77	49.31	62.70	34.02	54.18	-17.66
F27	34.61	42.02	76.77	6.88	-8.92	77.90	18.86	105.04	33.46
F28	29.92	77.48	72.74	38.13	33.67	77.57	-17.63	105.66	26.92
F29	-7.12	41.10	34.07	-11.99	4.49	36.05	43.83	81.85	21.53
F30	8.45	33.93	26.33	-10.21	-0.83	36.49	42.54	82.35	4.34
F31	9.24	46.36	40.90	31.22	-15.56	59.98	42.95	87.21	20.79
F32	7.89	37.58	22.47	19.05	7.61	44.82	36.76	74.17	2.37

Los resultados respecto a las ganancias para todas las fracciones posibles de los factoriales fraccionados 2^{9-5} , 2^{9-4} , 2^{9-3} , 2^{9-2} , 2^{9-1} y el arreglo 2^9 completo se muestran en las tablas 8, 9, 10, 11, 12 y 13 respectivamente.

Tabla 9: Ganancias utilizando las 16 fracciones posibles del arreglo 2^{9-4}

F 2^{9-4}	A	B	C	D	E	F	G	H	I
F1	39.50	101.65	91.09	-43.42	7.55	85.72	65.44	137.34	-2.32
F2	-18.50	128.56	118.35	-2.06	-1.58	131.32	74.62	146.73	51.13
F3	62.28	88.92	143.93	17.58	-17.96	118.38	57.73	158.90	48.08
F4	50.310	110.80	136.99	9.22	-5.80	123.43	93.28	74.85	4.74
F5	46.21	107.30	145.53	-24.80	-18.10	118.78	100.52	81.46	7.36
F6	59.85	92.26	139.14	-26.80	-20.28	113.13	47.21	144.74	40.51
F7	13.87	111.93	101.33	10.75	-26.49	101.80	39.89	114.70	35.95
F8	69.40	127.29	117.24	31.91	-8.58	105.42	87.25	163.74	7.78
F9	79.84	146.96	135.14	-9.63	50.65	118.50	92.18	167.60	43.39
F10	-12.06	118.37	108.26	43.81	46.99	125.93	53.81	131.02	15.88
F11	63.17	82.02	146.88	14.75	46.76	116.56	66.41	152.30	0.99
F12	43.35	108.68	137.72	-15.87	20.82	123.42	95.04	80.79	9.40
F13	46.32	111.47	140.33	3.53	16.73	118.68	98.52	91.36	14.62
F14	74.08	107.33	153.09	-3.76	64.59	136.53	26.70	163.03	25.10
F15	8.11	102.48	90.47	-38.73	33.03	102.19	38.82	116.93	24.47
F16	39.50	101.65	91.09	-43.42	7.55	85.72	65.44	137.34	-2.33

Tabla 10: Ganancias utilizando las 8 fracciones posibles del arreglo 2^{9-3}

2^{9-3}	A	B	C	D	E	F	G	H	I
F1	78.54	194.93	228.98	-39.86	-8.32	218.22	117.65	212.89	-4.31
F2	70.26	219.26	245.14	-10.09	18.08	220.29	117.35	215.48	17.70
F3	136.23	222.31	281.49	-22.28	29.03	256.36	152.18	256.27	41.19
F4	101.43	219.13	246.75	-5.52	44.11	219.54	165.71	289.13	58.10
F5	52.17	251.96	266.91	9.86	45.87	251.74	131.22	232.17	49.60
F6	74.75	211.74	235.13	0.88	-15.85	210.26	145.17	272.02	54.24
F7	119.22	214.34	265.86	1.47	59.03	252.37	138.34	317.06	74.67
F8	78.54	194.93	228.98	-39.86	-8.32	218.22	117.65	212.89	-4.31

Tabla 11: Ganancias utilizando las 4 fracciones posibles del arreglo 2^{9-2}

2^{9-2}	A	B	C	D	E	F	G	H	I
F1	168.77	411.43	479.45	-52.03	-18.34	430.18	257.71	496.05	50.60
F2	187.76	423.09	490.63	-23.62	63.41	451.97	275.31	504.23	83.09
F3	159.69	475.83	530.60	-13.91	45.97	485.49	315.43	547.52	114.40
F4	168.77	411.43	479.45	-52.03	-18.34	430.18	257.71	496.05	50.60

Tabla 12: Ganancias utilizando las 2 fracciones posibles del arreglo 2^{9-1}

2^{9-1}	A	B	C	D	E	F	G	H	I
F1	321.30	852.60	976.70	-51.90	59.30	889.80	543.50	1007.00	108.50
F2	365.20	919.40	1043.10	-27.60	121.20	953.00	573.70	1076.50	221.90

Tabla 13: Ganancias utilizando el arreglo 2^9 completo

	A	B	C	D	E	F	G	H	I
Completo	686.50	1772.00	2019.80	-	180.50	1842.80	1117.20	2083.50	330.50
				79.40					

RESULTADOS

Después de realizadas las pruebas para todos los arreglos se pudo ver que tan solo dos de ellos llevaron a las mismas conclusiones a las que se llegó en el artículo de referencia discriminando a los factores A, D y E por tener una ganancias negativas, una de ellas es una fracción de las 16 posibles del factorial fraccionado 2^{9-4} (F2, Generadores: F = BCDE, G = -ACDE, H = -ABDE, J = -ABCE) y la otra es una de las 32 fracciones posibles del factorial fraccionado 2^{9-5} (F13, Generadores: E = -ABC, F = -BCD, G = ACD, H = ABD, I = -ABCD), en todos los arreglos restantes los resultados fueron diferentes a los del trabajo de referencia, lo cual se puede verificar en las tablas de la 8 a la 13. Al revisar los resultados de los 63 arreglos se puede apreciar que de los factores rechazados utilizando el AO L_{12} reportados por Cudney et al. (2007), el factor A solo tuvo ganancia negativa en 6 ocasiones, el factor D en 39 y el factor E en 25 de ellas, mientras que el factor I que en el estudio original permaneció como variable importante tuvo ganancia negativa en 13 ocasiones, es decir, 7 ocasiones más que el factor A, a su vez el factor G tuvo ganancia negativa en 3 ocasiones, mientras que en los factores B, C, F y H la ganancia siempre fue positiva, por lo que quizá estos últimos debieran ser los únicos factores a considerar para posteriores análisis.

CONCLUSIONES

Después de realizar los cálculos para las 62 fracciones posibles y el arreglo 2^9 completo, se puede concluir que cuando se utiliza el SMT, la selección de factores importantes a considerar en un estudio dependerá del arreglo que se elija para alojarlos, por lo que se puede decir que el SMT no es robusto a los diferentes niveles de fraccionado, ni a las diferentes fracciones (diferentes generadores) dentro de cada nivel, lo cual es consistente con lo reportado por Rodríguez et al. (2010). El hecho de que el resultado dependa del arreglo utilizado es totalmente indeseable, ya que pudiera llegarse a conclusiones diferentes, es decir, a incluir o rechazar diferentes factores, debido al arreglo que se haya seleccionado dependiendo de los recursos con los que se cuente para el análisis de los datos. Incluso las conclusiones pudieran diferir utilizando arreglos con la misma cantidad de corridas pero contruidos con diferentes generadores, lo que demandaría la misma cantidad de recursos en la fase del procesamiento de datos. Bovas y Asokan (2003) también reportaron inconsistencias en la identificación de variables importantes usando el SMT cuando los factores se arreglaban de diferente manera en las columnas de la matriz de diseño y comentan que esas diferencias se pueden deber a que la selección de las variables se basa simplemente en sí sus efectos (ganancias) son positivos y no en una prueba estadística formal. Por los detalles antes mencionados se recomienda tener cautela con las conclusiones alcanzadas cuando se trabaja con el Sistema Mahalanobis - Taguchi.

NOMENCLATURA

Símbolo	Descripción
DM	Distancia de Mahalanobis
Z_i	Vector estandarizado obtenido de los valores estandarizados de X_i ($i=1, \dots, k$)
k	Número de características/ variables
C^{-1}	Matriz de correlación inversa
T	Transpuesta de un vector
S_i	Desviación estándar de la i-ésima característica
m_i	Media de la i-ésima característica
X_i	Valor de la i-ésima característica
SR^+	Representa la razón señal a ruido de incluir la variable
SR^-	Representa la razón señal a ruido de no incluir la variable

REFERENCIAS

- Antony, J. y Antony, F. J., *Teaching the Taguchi methods to industrial engineers*, Work Study, 50(4), 141-149, (2001).
- Báez, Y. A. y otros tres autores, *Aplicación de seis sigma y los métodos Taguchi para el incremento de la resistencia a la prueba de jalón de un diodo emisor de luz*, Información Tecnológica, 21(1), 63-76 (2010).
- Bovas, A. y Asokan, M. V., *Discussion of the Mahalanobis-Taguchi system*, Technometrics, 22-24, (2003).
- Cudney, E., Drain, D. P. y Naresh, S., *A comparison of the Mahalanobis-Taguchi system to a standard statistical method for defect detection*, Journal of Industrial and Systems Engineering, 250-258, (2009).
- Cudney, E., Paryani, K. y Ragsdell, k., *Applying the Mahalanobis-Taguchi system to vehicle ride*, Journal of Industrial and Systems Engineering, 1(3), 251-259, (2007).
- Cudney, E., Paryani, K. y Ragsdell, K., *Identifying useful variables for vehicle braking using the adjoint matrix approach to the Mahalanobis-Taguchi system*, Journal of Industrial and System Engineering, 281-292, (2008).
- Jain, A., Duin, R. y Mao, J., *Statistical pattern recognition: A review*. Transaction on Pattern Analysis and Machine Intelligence, 4-37, (2000).
- Jugulum, R. y Samuel, P., *Design for lean six sigma: A holistic approach to design and innovation*, John Wiley and Sons, 225 – 231, (2008).
- Rodríguez, M. A. y otros seis autores, *A robustness evaluation of Mahalanobis-Taguchi system using fractional factorial designs*, International Journal of Industrial Engineering, Special Issue, 479-486, (2010).
- Taguchi, G. y Jugulum, R., *New trends in multivariate diagnosis*, Indian Journal of Statistics, 233-248, (2000).
- Taguchi, G. y Jugulum, R., *The Mahalanobis-Taguchi strategy: A pattern technology system*, John Wiley and Sons, New York, (2002).
- Taguchi, G., Chowdhury, S. y Wu, Y., *Taguchi's quality engineering handbook*, Wiley, New Jersey, (2005).
- Taguchi, G., Chowdury, S. y Wu, Y., *The Mahalanobis-Taguchi system*, McGraw Hill, New York, (2001).
- Taguchi, G., Jugulum, R. y Taguchi, S., *Computer-based robust engineering: Essentials for DFSS*. Milwaukee, Wisconsin: Quality Press, (2005).
- Taguchi, G., *Taguchi methods research and development*, Dearbon: American Suppliers Institute Press, (1991).
- Wolberg, W. H., *Wisconsin breast cancer data base*, 1(8), 1991.
<http://www.uwplatt.edu/csse/courses/cs303/as/data/cancer/.html>. Acceso: 8 de enero (2009).
- Woodall, W. y otros cinco autores, *A review and analysis of the Mahalanobis-Taguchi system*, American Statistical Association, 45(1), 1-15, (2003).

