

## Identificación de desabasto de productos farmacéuticos utilizando un gráfico de control multivariante $T^2$ de Hotelling

M.I.I. Marco Antonio Morales Canul<sup>1</sup>, Dr. Sebastián Díaz de la Torre<sup>2</sup>, Dr. T.A. José Claudio Cenobio Méndez García<sup>3</sup>

**Resumen:** Realizar auditorías en campo resultan ser tareas difíciles de realizar, además de costosas. En esta investigación se analizan 215 farmacias que mercan 15 diferentes productos. El objetivo es identificar la variación del desabasto, utilizando el gráfico de control multivariante  $T^2$  de Hotelling. Este método nos permite analizar grandes volúmenes de datos e identificar visualmente el comportamiento de múltiples variables de manera sencilla, eficiente y económica. Aquí se identifica el comportamiento de la falta de medicamentos en 5 cadenas comerciales.

**Palabras clave:** Grafico de control de multivariante,  $T^2$  de Hotelling, desabasto, farmacias, R Project.

### Introducción

Son casi nulas las empresas que aplican estadística multivariante en la solución de problemas. Consideramos como ejemplo a una empresa farmacéutica que desconoce el comportamiento de desabasto en su línea de distribución; tiene continuas quejas de sus clientes que le amenazan con ya no comprarle por no tener productos. La compañía ha realizado inspecciones físicas en cada lugar donde se ubican las farmacias, sin éxito, pues les fue difícil contar los miles de piezas distribuidas en cada una de ellas además de obstruir la operación de sus clientes. Dicha situación es muy recurrente en micro y pequeñas empresas que no cuentan con algún planificador de requerimientos o con un enlace en la red de datos, la llamada nube.

Usar control multivariante conlleva ventajas sobre otros métodos tradicionales de inspección. En tanto, analizar variables una por una para un contexto general, resulta ser un trabajo arduo que no contribuye a tener información simplificada; utilizar gráficos de control multivariante, en cambio, permite procesar datos masivamente que monitorean secuencialmente la relación entre múltiples procesos, aportando información veraz con imágenes visuales sencillas de entender.

Aquí se presenta la creación de un gráfico de control multivariante  $T^2$  de Hotelling a partir de una base de datos que contiene información de 15 productos farmacéuticos en 215 farmacias que son los puntos de venta, identificando plenamente el comportamiento general de desabasto de todas sus cadenas comerciales.

### Marco Teórico

#### Ecuación $T^2$ de Hotelling

De forma generalizada, el algoritmo matemático que soporta el grafico de control multivariante  $T^2$  de Hotelling (Mason & Young, 2002) es:

$$T^2 = n(x - \bar{x})' \Sigma^{-1} (x - \bar{x}) \quad \text{Ecuación 1}$$

Donde:

$n$  es el número de observaciones por cada muestra (Si fuese solo una observación por registro este valdría 1).

$x$  es el vector de muestras.

$\bar{x}$  es el vector de medias muestrales que representa una medida del centro del proceso.

$(x - \bar{x})$  Es la diferencia con respecto a las medias muestrales individuales.

$(x - \bar{x})'$  Es la transpuesta de la matriz  $(x - \bar{x})$ .

$\Sigma^{-1}$  es la matriz inversa de la varianza-covarianza.

Se define la matriz de varianzas - covarianzas como:

$$\Sigma = E[(x - E[x])(x - E[x])'] \quad \text{Ecuación 2}$$

<sup>1</sup> El M.I.I. Marco Antonio Morales Canul es alumno de doctorado en tecnología avanzada en el Centro de Investigación e Innovación Tecnológica del Instituto Politécnico Nacional en Ciudad de México, Consultor empresarial y Profesor de asignatura. [mmoralesc8400@alumno.ipn](mailto:mmoralesc8400@alumno.ipn). (autor correspondiente)

<sup>2</sup> El Dr. Sebastián Díaz de la Torre tiene Doctorado Honoris Causa por la Brno University Technology, de la República Checa, actual director del Centro de Investigación e Innovación Tecnológica del Instituto Politécnico Nacional en Ciudad de México, [sediazt@ipn.mx](mailto:sediazt@ipn.mx)

<sup>3</sup> El Dr. T.A. José Claudio Cenobio Méndez García es Profesor del Centro de Investigación e Innovación Tecnológica del Instituto Politécnico Nacional en Ciudad de México, [jcmendez@ipn.mx](mailto:jcmendez@ipn.mx)

Donde:

$x$  es el vector de muestras.

$E$  es la suma de cada operación entre corchetes dividida por el número de elementos del vector que lo componen. Es decir, en términos prácticos es el promedio.

$(x - E[x])'$  es solo la transpuesta de la matriz  $(x - E[x])$ .

En ingeniería, esta es la nomenclatura que ocupamos, aunque cabe aclarar que en algebra lineal esta difiere; por ejemplo, en el sentido más puro matemático, es característico utilizar letras mayúsculas para denotar las matrices.

#### *Gráfico de control $T^2$ de Hotelling*

El gráfico de control  $T^2$  de Hotelling da seguimiento simultáneo a dos o más variables relacionadas con una secuencia de tiempo; identifica el comportamiento atípico en el periodo exacto y lo hace de manera gráfica.

La importancia radica en que examina simultáneamente la interacción de la varianza y la covarianza de todas las variables del proceso.

Como se entiende, el gráfico se puede aplicar en situaciones reales, puesto que ofrece una ayuda para comparar varias unidades de medición tanto físicas como químicas u otras.

$T^2$  permite contrastar proporcionalmente todas las variables en un solo gráfico, sin importar la cantidad de estas, considerando las covarianzas entre variables.

El gráfico de control  $T^2$  identifica el punto exacto donde hay un punto atípico, sin embargo, tiene el inconveniente de no reconocer cuál de las variables fue la que lo ocasionó la diferencia.

#### *Límite de control $T^2$ de Hotelling*

El límite más exacto es considerar la distribución Fisher-Snedecor, sin embargo, Peña (Peña, 2002) descubrió que, al aumentar el número de muestras de las variables, la probabilidad de Fisher se asimila con la de Chi-cuadrado. Con lo cual, se simplifica aún más el cálculo del límite de control correspondiente para  $T^2$  de Hotelling.

Por lo tanto, aquí recomendamos utilizar el estadístico de prueba  $\chi^2$  Chi cuadrado.

$$LSC = \chi^2_{\alpha,p} \quad \text{Ecuación 3}$$

De donde:

$LSC$  es el límite de control superior y único.

$\chi^2$  es el estadístico de prueba de la distribución Chi cuadrado.

Donde  $\alpha$  es la probabilidad o nivel de significancia.

$p$  es el número de variables.

Esta ecuación puede ser utilizada para muestras con 30 o más elementos, y se puede usar con  $3\sigma$  (desviación estándar) para monitorear el proceso como estimación inicial.

### **Desarrollo**

#### *Metodología General*

Para el desarrollo, se siguieron los pasos que se muestran a continuación:

1. Selección de variables
2. Colección de datos en campo
3. Revisión previa de datos
4. Creación del gráfico de control  $T^2$  multivariado de Hotelling
5. Identificación de valores atípicos (Datos por encima de la línea de control LSC)

En la siguiente sección, se describen los primeros cuatro pasos detallados, y en los Resultados y Análisis se identificaron los valores atípicos (el quinto paso).

La empresa en cuestión deseó permanecer en el anonimato.

#### *1. Selección de variables*

La farmacéutica tiene 215 farmacias distribuidas entre 5 cadenas, como se muestra en la siguiente tabla:

	<i>Cadena</i>	<i>Cantidad Farmacias</i>
1	Ftodo	40
2	Snbrns	30
3	Spablo	87
4	Sprma	28
5	Wmart	30
	<b>Total</b>	<b>215</b>

Tabla 1. Cadenas y cantidad de farmacias con que cuentan

Los productos que maneja la farmacéutica son 15; estos productos se desean que tengan existencia en cada una de las 215 farmacias para ser vendidos.

Las claves de los productos son: A100, EEB, EG, G50\_1, G100\_1, G100\_4, Ple, R, R20, S30, S60, S90, T100\_30, TEB, TP.

Anteriormente, la empresa intentó hacer varios inventarios de forma tradicional, contra las existencias físicas de producto por cada farmacia, lo cual aparte de ser mucho trabajo, incomodaba a los propietarios.

## 2. Colección de datos en campo

Para esta investigación, un grupo de 10 personas fueron contratadas para hacer investigación en campo.

La solución fue registrar los datos de cada farmacia en una base binaria que posea las siguientes características:



<b>Existencia en cada farmacia</b>		<b>Se registra en la base de datos</b>
	No hay ninguna pieza del producto	1
	Hay al menos una o más piezas del Producto	0 (cero)

Tabla 2. Consideraciones a tomar para el registro del desabasto en la base de datos

Es decir, se visualiza la inexistencia de cada producto: si no había ninguno se captura 1, y si por lo menos del producto había 1, se pone cero. Como se puede observar, la base de datos permite identificar de inmediato los productos que tienen desabasto o nula existencia, pues es fácil de identificar por la asignación del número 1.

La Figura 1 nos muestra el número de farmacias visitadas (censadas) por el grupo de investigadores, como se observa en la primera visita se censaron 210 farmacias, en la segunda visita 212, hasta llegar a la última que fue la visita número 8 con 199 farmacias inspeccionadas.

En la misma Figura 1 se observa que la visita 5 tiene el mínimo de farmacias auditadas con 187 y la máxima la visita 3 con 213 farmacias examinadas.

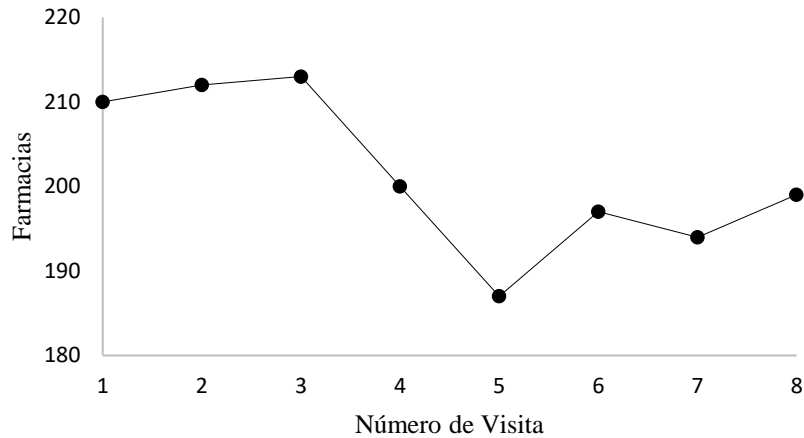


Figura 1. Cantidad de farmacias auditadas por cada visita

Número de Visita	1	2	3	4	5	6	7	8	Total
Farmacias auditadas	210	212	213	200	187	197	194	199	<b>1612</b>

Tabla 3. Cantidad de Farmacias auditadas por cada visita (Incluyen varias cadenas)

La Figura 2 representa la integración de farmacias inspeccionadas por las 8 visitas, agrupadas por cadena comercial. Las unidades que más se trabajaron fueron SPablo y Ftodo.

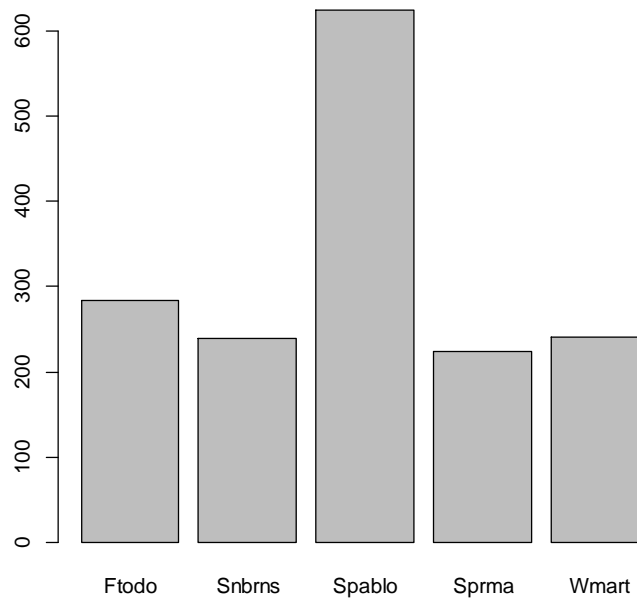


Figura 2. Cantidad de farmacias auditadas por cadena en todas las visitas

### 3. Revisión previa de datos

En este paso, es muy importante que antes de procesar cualquier base de datos se haga una revisión preliminar. La Tabla 4 resume algunas características que se consideraron en este proceso:

✓	Registros con omisión de información
✓	Duplicación de registros.
✓	Búsqueda de datos ilógicos (Extremadamente pequeños o grandes, comparados con sus homólogos de clase).
✓	Detección de registros incorrectos, ocasionados entre otras causas por distracción, cansancio, mala calibración de los instrumentos, errores de captura, Incluso en situaciones críticas por negligencia.
✓	Validación de registros excluidos por anomalía, pero sin afectar los que realmente son puntos atípicos generados por el manejo natural del proceso.

Tabla 4. Puntos por considerar en la revisión previa de datos

Aquí se solicitó la asistencia de las personas que realizaron el inventario en campo, para validar que la información de sus registros fuera correcta.

La captura final, posterior a la revisión, se ejemplifica en la Tabla 5.

Reg*	Clave	Cadena	Visita	PRODUCTO														
				A100	EEB	EG	TEB	TP	G100_4	G100_1	G50_1	T100_30	R20	R	S30	S60	S90	Ple
1	FT041	Ftodo	1	1	0	1	1	1	1	0	1	0	0	1	1	1	0	0
2	FT022	Ftodo	1	0	0	0	0	0	1	1	1	0	0	0	0	1	1	1
3	FT034	Ftodo	1	0	1	0	1	0	1	0	1	0	0	0	0	1	1	0
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
1612	WT012	Wmart	8	1	0	0	0	0	0	0	0	0	1	1	0	0	0	0

\* Número de registro cronológico.

Tabla 5. Ejemplo de base de datos capturada

En la Tabla 6 se indica la suma de desabasto de medicamento.

Las auditorías a las farmacias se agruparon con su respectiva cadena por visita.

El desabasto manifiesta la suma de cada uno de los 15 productos diferentes de la empresa que al ser censados no tuvieron ni una sola existencia.

Se reitera que cada farmacia debería tener por lo menos uno de los 15 productos diferentes de la farmacéutica.

No se deben confundir los datos de la Tabla 6 cuyo método utilizado fue el manifestado en la Tabla 2, que son muy diferentes con los inventarios tradicionales de contar las existencias.

También es importante aclarar que en la Figura 2 son cantidad de farmacias y en la Tabla 6 son productos.

Visita	Cadena				
	Ftodo	Snbrns	Spablo	Sprma	Wmart
1	222	238	344	268	227
2	225	244	373	266	230
3	219	250	463	275	235
4	207	248	423	262	224
5	153	247	416	273	224
6	168	257	352	277	228
7	157	250	259	279	242
8	157	255	259	277	236
<b>total</b>	<b>1508</b>	<b>1989</b>	<b>2889</b>	<b>2177</b>	<b>1846</b>

Tabla 6. Productos con cero existencias por cadena

El total de productos que no tenían ninguna existencia es la suma de  $1508 + 1989 + 2889 + 2177 + 1846 = 10,409$ .

Sí se usa un análisis tradicional, se darían conjeturas erróneas, ya que en la Figura 2 se aprecia que SPablo tendría más desabasto, sin embargo, se debe considerar que esta cadena consta de 80 farmacias contra Sprama que tiene solo 28 (Tabla 1), además las visitas se realizaron con diferente número de farmacias examinadas (Tabla 3).

#### 4. Creación del gráfico de control $T^2$ multivariado de Hotelling

Obtener información más fidedigna, implicaría hacer centenares de gráficos, tratando de encontrar relaciones entre todas ellas, sin la seguridad de poder conseguirlo en su conjunto. Si deseamos saber si es proporcional con su número de tiendas por cadena comercial, productos con desabasto y visitas, todo en una única relación, entonces, necesitamos hacer un gráfico de control multivariante  $T^2$  de Hotelling por que involucra más de una variable.

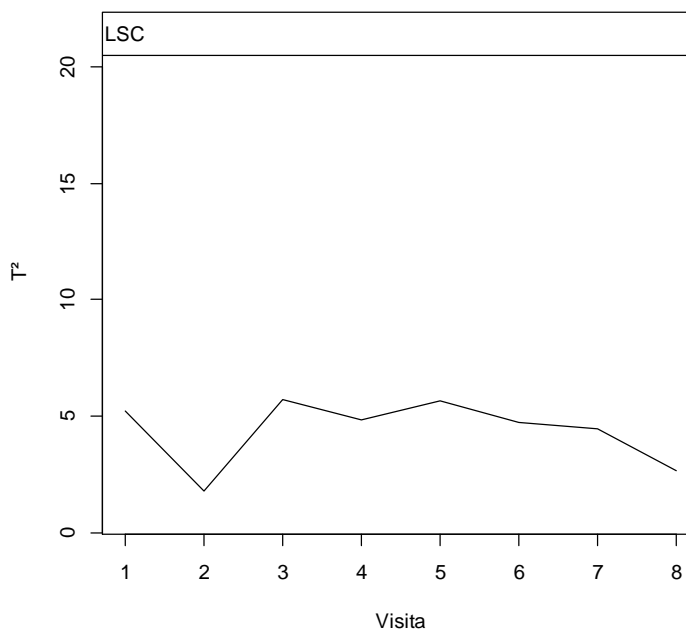


Figura 3. Gráfico de control multivariante  $T^2$  de Hotelling

Para decirlo de manera más sencilla, deseamos saber si se está favoreciendo específicamente a alguna cadena o cuál de ellas tiene un mayor desabasto. En la Figura 3 se presenta el gráfico de control multivariante  $T^2$  de Hotelling tomando como base la Tabla 6.

Para construir la Figura 3 se utilizó el software R Project, haciendo un algoritmo matemático que realice las operaciones matriciales de  $T^2 = n(x - \bar{x})' \Sigma^{-1} (x - \bar{x})$ , considerando un nivel de significancia  $\alpha=0.001$  y G.L.= grados de libertad = 5 = Número de cadenas de farmacias. Nos da un límite superior de control y único  $\chi^2$  de 22.46.

En la Figura 3, cada punto representa la variación en conjunto del desabasto de los medicamentos, de todas las farmacias auditadas por cada visita realizada. Además, cada cadena está estandarizada para poder compararlas con las demás.

El algoritmo matemático creado para obtener los valores individuales de  $T^2$  y que se puede reproducir para cualquier otra matriz es:

```
# a= matriz con valores numéricos y encabezados en la primera fila
# (Sin considerar la primera columna de registro)
library(readr)
library(tidyverse)

A<-matrix(rep(matrix(c(apply(a,2, mean))), nrow(a)), nrow=nrow(a),
ncol=ncol(a), byrow=TRUE)
M<-data.matrix(a-A)
I<-data.matrix(solve(cov(a)))
T²<-as.numeric(diag(M%*%I%*%t(M)))
```

```
T2

alfa<-0.001
GL<-ncol(a)
LSC<-qchisq(alfa,GL,lower.tail=F)
LSC

H<-a%>%
mutate(Hotelling=T2,
Num = 1:nrow(a))%>%
filter(Hotelling>LSC)
write.csv(H, file="outliers.csv")
view(H)
```

### Pruebas y Resultados

#### 5. Identificación de valores atípicos (Datos por encima de la línea de control LSC)

En la Figura 3, el eje de las ordenadas de cada punto  $T^2$  posee 3 características descritas a continuación:

- Representa la variación en conjunto del desabasto de todos los productos cronológicamente por visita.
- Cada cadena comercial esta estandarizada para compararlas con las demás variables, debido a la aplicación de la inversa de su matriz de varianza – covarianza.
- Considera la variación entre las cadenas, al realizar el producto de matrices de la diferencia con respecto a sus medias de desabasto.

En el eje de las abscisas están contempladas cronológicamente las visitas donde cada una de ellas contiene un número diferente de farmacias inspeccionadas.

Para considerar un punto atípico, los puntos deberían estar por arriba de la línea LSC; como ningún punto cumple con esa condición, esto implica que ninguna cadena tiene un mayor desabasto con respecto a las demás.

El uso tradicional de graficas con una sola variable puede ocasionar errores en la interpretación de los datos, por ejemplo, aquí se aprecia que la cadena Spablo es posiblemente la que tiene mayor desabasto, sin embargo, no se contemplan otras variables como número de tiendas de la cadena y las veces que fueron auditadas.

Los resultados de la investigación de campo incluyen el análisis estadístico multivariante para identificar 10,409 productos que no tenían ni una sola unidad disponible (desabasto) realizada durante la auditoria de 1,612 farmacias en la inspección de 8 visitas. Se obtuvo un gráfico  $T^2$  a partir del método propuesto, observando que en ninguna de las 8 visitas existe una cadena farmacéutica que no tenga desabasto. Es decir, el problema es generalizado y que la empresa tiene un severo problema de abastecimiento de su producto en la totalidad de sus 5 cadenas comerciales. No hay prueba estadística que confirme que alguna cadena tiene preferencia o que no tenga un desabasto similar a las demás.

Por medio del análisis del gráfico  $T^2$ , la compañía farmacéutica logró identificar y focalizar plenamente la situación de la falta de producto. Se le facilitó la toma de decisiones fundadas en evidencias cuantitativas.

### Conclusiones

La utilización del gráfico de control multivariante  $T^2$  de Hotelling simplifica, esclarece y hace más certero el estudio de los datos, aquí de una manera muy significativa se simplifico a una sola gráfica.

$T^2$  de Hotelling permite contrastar proporcionalmente todas las variables en un solo gráfico, sin importar la cantidad de estas, considerando las covarianzas entre variables.

También, identifica el punto exacto donde hay un punto atípico, pero tiene el inconveniente de no reconocer cuál de las variables fue la que lo ocasionó.

Para casos similares a este, muchas veces se emplean distintas técnicas de análisis multivariante; esto sería un tema muy interesante para la publicación de otro artículo.

Para valorar lo estratégico de utilizar  $T^2$  lo compararemos con el gráfico de control Shewart que es univariante. Al usar Shewart en un análisis global, se necesita hacer una gráfica por cada variable, lo cual implica mucho trabajo y si quisiéramos hacer una comparación entre variables, es prácticamente imposible o muy difícil.

Esta investigación es la punta del iceberg en la aplicación de análisis multivariante ya que se pueden usar otras técnicas. En el presente caso se podrían usar para encontrar el producto que tiene el mayor desabasto, la zona geográfica donde se concentra o encontrar indicadores que permitan evaluar una tendencia en el tiempo.

Todos los datos son diferentes; uno de los grandes méritos de utilizar cualquier método multivariante es que permite simplificar los análisis. Es tan poderosa la herramienta que resulta difícil creer que en una gráfica tan sencilla como la de la Figura 3 pueda tener concentrado la información de miles de datos.

Los procesos de manufactura y comercialización son sistemas abiertos, es decir, el resultado de un proceso se debe a la interacción de múltiples agentes tanto externos como internos. Los registros pueden ser generados en cantidades muy grandes. Hoy es la era denominada “Revolución del big data” (Puertas, Pujol Vila, & Seguí, 2017), porque contamos con mayores capacidades de procesamiento y asistencia de softwares, es por eso la imperiosa necesidad de utilizar métodos de análisis más completos que sean multivariantes.

Podríamos intuir que hay un inmenso campo aun por explorar, que trascienden a la solución de problemas complejos.

### *Limitaciones*

Aunque la interpretación del gráfico  $T^2$  de Hotelling es sencilla, la programación del algoritmo de algebra lineal que se hace en el software de R Project puede ser complejo y probablemente precise de un científico de datos.

El estudio en campo se limitó al presupuesto para salarios de personal operativo y de supervisión, además de gastos de telefonía y transportación para los mismos.

Otra de las limitaciones, fue la estandarización de trabajo en el registro de información, dado que se hizo todo de forma manual, se dependió de la eficiencia y efectividad de cada auditor.

Se recomienda utilizar medios electrónicos como la web o internet, incluso de tener los recursos, implementar un programa informático de Gestión de la Relación con el Cliente.

### **Referencias**

- Mason, R., & Young, J. (2002). *Multivariate statistical process control with industrial applications*. Houston: ASA SIAM.  
Peña, D. (2002). *Análisis de datos multivariantes*. España: McGraw-Hill.  
Puertas, E., Pujol Vila, O., & Seguí, S. (2017). *El poder de los datos*. Barcelona: RBA Ibérica.

### **Notas Biográficas**

El M.II. Marco Antonio Morales Canul, es Ingeniero Mecánico egresado de la ESIME del Instituto Politécnico Nacional, Maestro en Ingeniería Industrial por el Instituto Tecnológico de Celaya; actualmente es alumno de doctorado en tecnología avanzada en el Centro de Investigación e Innovación Tecnológica del Instituto Politécnico Nacional en Ciudad de México, ha publicado artículos en la revista Academia Journals. Ha asesorado a micro y pequeñas empresas, así como transnacionales y es profesor de asignatura en el Tecnológico de Estudios Superiores de Ecatepec.

El Dr. Sebastián Díaz de la Torre es Ingeniero Químico por la Universidad Autónoma de Zacatecas. Tiene su Maestría en Ingeniería Metalúrgica en la ESIQIE del IPN. D.F., México y Doctorado en Ingeniería y Ciencia de Materiales del Departamento de Ciencias de la Energía e Ingeniería de la Universidad de Kyoto, Japón. Actualmente es el Director del Centro de Investigación e Innovación Tecnológica del IPN en Ciudad de México, Maestro Decano, asesor externo de la Comisión de Ciencia y Tecnología del Senado de la República, y ha escrito múltiples publicaciones científicas en la editorial Elsevier entre otras. Tiene su Doctorado Honoris Causa por la La Brno University Technology, de la República Checa.

El Dr. T.A. José Claudio Cenobio Méndez García tiene un Doctorado en Tecnología Avanzada egresado del CIITEC del IPN, tiene Maestría en Enseñanza Superior por la UNAM, Maestría en Administración de Empresas por el ITESM, es Ingeniero Metalúrgico por la ESIQIE del IPN, y docente del CIITEC y UPIICSA del IPN; ha escrito diferentes libros y publicado en diferentes revistas científicas.