

EXCEL APLICADO EN ESTADÍSTICA: REGRESIÓN LINEAL Y POLINOMIAL

M. A. Lerma García¹
P. L. Lerma García²
L. S. Lerma Rojas³

RESUMEN

Los docentes buscan que sus clases sean útiles, amenas y divertidas, donde el aprendiz logre un esquema de aprendizaje sustantivo y de algún modo efectivo, mismo que permita practicar los conocimientos dispensados en el aula, así, logran en buena medida la asimilación del material necesario para demostrar el aprendizaje y lograr una aprobación sino rotunda al menos aceptable y, a partir de éste, los aprendices podrán incursionar en técnicas y modelos más complejos, esto implica precisamente, trabajar preliminarmente para diseñar y desarrollar algunos métodos y técnicas conducentes a tal fin, generando herramientas que cotidianamente facilitan el estudio y el trabajo. De tal forma, concebir un procedimiento o un recorrido que admita sin problema obtener la mayor comprensión con el mínimo de los empeños y de tiempo, la objeción es, no impera esa aptitud y, con los desenlaces convenientes. Se muestra un proceso de cálculo, para un análisis básico de la estimación de la ecuación de regresión parabólica, con similitud para las ecuaciones de regresión potencial y lineal múltiple, determinando los tabulados básicos, cálculo de inversa por método de Gauss-Jordan y regla de Kramer-Raho, presentando el modelo de regresión estimado. Se trata de explicar los métodos tradicionales empleando Excel de Microsoft Office, con su utilidad como herramienta de enseñanza-aprendizaje, obteniendo resultados superiores al método tradicional (Manual) generalizado.

ANTECEDENTES

La regresión lineal simple es un tema bastante interesante, se considera que su grado de dificultad puede parecer simple para algunos docentes, en cambio para los estudiantes o aprendices puede conllevar ciertos aspectos de desarrollo y análisis que le dan poco atractivo y pueden llegar a parecer complejos y complicados como cualquier otro tema que requiere un análisis matemático.

Es un tema selecto para varios programas de ingeniería, además de ciencias económicas, sociales, medicina y cualquier programa de estudios profesionales y de posgrado que requieran un modelo de predicción para estimaciones futuras, pasadas y posibles correlaciones.

En la investigación de todo proyecto, se parte de un potencial inconveniente, problemática o interrogante, en este caso específico ¿Para qué es obligatorio adecuar la habilidad de enseñanza-aprendizaje a la Formación Basada en Competencias?, pretendiendo como objetivo allegar al aprendiz, un proceso de aprender haciendo, conllevando la práctica en la secuencia tradicional y aplicación-ejercitación vía Excel.

Se toman dos programas de ingenierías: de Industrial y de Gestión Empresarial. Los programas son tan solo dos ejemplos de entre muchos más programas existentes en la

¹Catedrática del Tecnológico Nacional de México, campus Durango. Departamento de Metal-Mecánica. mara.lerma@itdurango.edu.mx (Autora contacto)

²Catedrático del Tecnológico Nacional de México, campus Durango. Departamento de Eléctrica-Electrónica. pedro.lerma@itdurango.edu.mx

³Investigador del Tecnológico Nacional de México, campus Durango. Departamento de Ciencias Económico Administrativas. llerma@itdurango.edu.mx

educación-formación pública y privada de nivel superior de México. Dichos programas que se citan: Estadística Inferencial II, de la Ingeniería en Gestión Empresarial (Tecnológico Nacional de México [TecNM], 2010) y Estadística Inferencial I, del así, denominado, modelo educativo basado en competencias 2010 (TecNM, 2016), en ambos programas los temas de Regresión lineal simple se encuentran ubicados en la unidad temática 1 y 5 respectivamente.

De igual forma se enuncia, las competencias específicas a desarrollar en ambos programas (TecNM, 2010, 2016):

- Aplica, desarrolla y analiza las técnicas de regresión lineal simple para hacer predicciones de sucesos futuros en el ramo empresarial.
- Identifica y aplica los conceptos básicos del modelo de regresión lineal simple, establece las condiciones para distinguir entre una regresión y una correlación e Identifica y aplica los conceptos básicos del modelo de regresión múltiple.

Se puede observar las grandes similitudes en dichas competencias. Cabe destacar que dichos programas incluyen sugerencias didácticas. Para uno de los programas se proponen las siguientes sugerencias didácticas (TecNM, 2016):

- Utiliza correctamente un modelo de regresión para propósitos de estimación y predicción.
- Comprende la importancia del análisis de regresión lineal simple y múltiple, y explique los conceptos generales.
- Aplica las pruebas de hipótesis para evaluar su calidad de ajuste.
- Diferencia entre regresión lineal simple y múltiple para tomar decisiones acerca de cuál modelo usar en determinada circunstancia.
- Comprende la importancia del análisis de regresión no-lineal y explique los conceptos generales.
- Aplica las pruebas de hipótesis para evaluar su calidad de ajuste.
- Utiliza software, para obtener una respuesta rápida y precisa en la generación de los parámetros de los modelos.

Para otro de los programas se brindan las sugerencias (TecNM, 2010):

- Identifica las variables dependientes e independientes para el análisis de regresión.
- Ajusta un modelo de regresión lineal que relacione una variable independiente (Controlable) y una variable dependiente (No-controlable).
- Analiza gráficas que permitan entender la relación existente entre las variables en consideración.
- Utiliza el análisis de regresión simple para estimar la relación entre las variables.
- Utiliza el coeficiente de correlación para medir el grado de relación lineal entre las variables.
- Obtiene el coeficiente de determinación para medir la fuerza de relación entre las dos variables.
- Interpreta los coeficientes de regresión.
- Desarrolla inferencias estadísticas para los coeficientes de regresión.

Nuevamente, se observa una gran similitud. Estas, analogías permiten proponer un proceso de cálculo en Excel de Microsoft Office para un análisis básico de la estimación de la ecuación de regresión polinomial de orden 2, con similitud para las ecuaciones de regresión polinomial y lineal múltiple, determinando los tabulados básicos, cálculo de inversa por método de $X = BA^{-1}$ (Spiegel y Stephens, 2011), presentando el modelo de regresión estimado.

Se trató de explicar los métodos tradicionales empleando una hoja electrónica como Excel de Microsoft Office, con su utilidad como herramienta de enseñanza-aprendizaje, logrando un aprendizaje y retención muy superior que el del método tradicional.

La selección de Excel como herramienta para el desarrollo asistido en los procedimientos y cálculos es su uso extendido y virtual instalación en equipos de distintas generaciones desde ordenadores relativamente antiguos hasta modernos dispositivos personales. Su invariabilidad de uso y técnicas para la captura y procesamiento de datos que no requiere una capacitación-preparación demasiado extensa. Además, se cuenta con un soporte en línea y en medios impresos suficientemente robusto para que los alumnos o aprendices interesados puedan perfilar y mejorar su conocimiento, aplicándolo a otras áreas de conocimiento durante su proceso de formación profesional.

Se cita de alguien que se escucha, pero, no se recuerda su nombre, sobre lo que debe saber todo profesionalista: “Todo profesionalista debe saber, tres aspectos dentro de su formación profesional: Inglés, Excel y decir si-no en el momento preciso”.

METODOLOGÍA

Se presentan los modelos que componen el análisis de regresión lineal simple por el método de mínimos cuadrados, el cociente de determinación. Además, se incorpora el método de regresión parabólica y, se aprovecha las virtudes del Excel para demostrar que el sistema de ecuaciones puede ser resuelto de una forma simple y sencilla con unos cuantos pasos extras en Excel, sobre todo habiendo aprendido haciéndolo manualmente.

Se aprovecha la forma simple de trabajar en Excel (Cardona, Rojas y Mesa, 2008) para desarrollar y resolver los métodos de ambas regresiones, ingresando los modelos y se explica cómo editar las ecuaciones de dichos modelos en Excel, con el propósito de determinar los parámetros de ambos modelos.

Finalmente, se comprueban los resultados de los procedimientos mediante funciones propias de Excel, además, de obtener las líneas de tendencia y los modelos de regresión.

La Herramienta o menaje ilustrativo es un desarrollo tabular o bosquejo representativo de simbolización similar a los macros de Excel, con apoyo de él, los aprendices se adaptan y asocian los planteamientos de los problemas a escenarios reales o vivenciales, valiéndose del raciocinio y de la lógica, lo más notable, la asociación de la palabra a una representación determinística; por ello, el aprendiz ubica los argumentos y lleva de esa manera cualquier hecho a una solución factible o representación de la realidad, sin omitir el estado del mismo; aprovechando para el fin didáctico el Excel y el Power Point de Microsoft Office con animaciones.

Aprovechando esta técnica, reiteradamente, se logra familiarizar al aprendiz, él deberá tener presente la imagen clara de que, los problemas están suscritos a una forma original de ocurrir de los acontecimientos de los variados eventos naturales, pudiéndose utilizar implícitamente con objetos o cosas y, es cuando el aprendiz distinguirá captando que cada objeto, cosa o persona están asociados entre sí.

El alcance dependerá del aprendiz y, del que enseña, agregando la continuidad, el perfeccionamiento de la teoría e integrará la intervención de los elementos del universo.

RESULTADOS

Se consideró un problema para desarrollar la propuesta:

Una empresa de mensajería de entrega puerta a puerta, con fin de mejorar la prestación de servicio desea establecer la relación que puede existir entre el tiempo empleado y la distancia recorrida para la entrega de un determinado producto, ejercite:

- Realice un diagrama de dispersión a partir de los datos obtenidos
- Determine el mejor modelo y su ecuación representativa que se ajuste a los datos

Los datos proporcionados para el problema se anotan en Excel y se designan las variables para comenzar con los cálculos de los valores necesarios para el desarrollo del modelo de regresión lineal (Véase Figura 1):

Regresion lineal simple por mínimos cuadrados					
Una empresa de mensajería de entrega puerta a puerta, con fin de mejorar la prestación de servicio desea establecer la relación que puede existir entre el tiempo empleado y la distancia recorrida para la entrega de un determinado producto					
*Realice un diagrama de dispersión a partir de los datos obtenidos					
*Determine la mejor ecuación que se ajuste a los datos					
Distancia (km)	Tiempo (Días)				
x	y		x*y	x^2	y^2
825	3.5		2887.5	680625	12.25
215	1		215	46225	1
1070	4		4280	1144900	16
550	2		1100	302500	4
480	1		480	230400	1
920	3		2760	846400	9
1350	4.5		6075	1822500	20.25
325	1.5		487.5	105625	2.25
670	3		2010	448900	9
1215	5		6075	1476225	25
Sumatorias	7620	28.5	26370	7104300	99.75
promedios	762	2.85			
promedios^2		8.1225		n=	10

Figura 1. Problema y datos tabulados. Fuente: Elaboración propia.

Las ecuaciones para el modelo de regresión lineal simple por el método de mínimos cuadrados son las ya conocidas y aplicadas (Véase Figura 2):

Modelo de ecuación lineal	Pendiente	Intercepto
$Y = a + bX$	$b = \frac{n \sum xy - \sum x \sum y}{n \sum x^2 - (\sum x)^2}$	$a = \frac{\sum y - b \sum x}{n}$
Error estándar del estimado	Varianza de la variable	Valor de correlación

$$Se = \sqrt{\frac{\sum y^2 - a \sum y - b \sum xy}{n - 2}} \quad \text{separable} \quad R^2 = 1 - \frac{Se^2}{S_y^2}$$

$$S_y^2 = \frac{\sum y^2}{n} - \bar{y}^2$$

	1350	4.5	6075	1822500	20.25
	325	1.5	487.5	105625	2.25
	670	3	2010	448900	9
	1215	5	6075	1476225	25
Sumatorias	7620	28.5	26370	7104300	99.75
promedios	762	2.85			
promedios^2		8.1225		n=	10
Pendiente			Intercepto		
$b = \frac{n \sum xy - \sum x \sum y}{n \sum x^2 - (\sum x)^2}$			$a = \frac{\sum y - b \sum x}{n}$		
b=			a=		
b= 0.003585132			a= 0.1181291		
Y=0.118+0.0036X					

Figura 2. Aplicación con datos en ecuaciones. Fuente: Elaboración propia.

Aplicando las ecuaciones de pendiente e intercepto, se obtiene el modelo de la ecuación lineal (Véase Figura 3). El modelo resultante:

$$Y = 0.118 + 0.0036X$$

	670	3	2010	448900	9
	1215	5	6075	1476225	25
Sumatorias	7620	28.5	26370	7104300	99.75
promedios	762	2.85			
promedios^2		8.1225		n=	10
Pendiente			Intercepto		
$b = \frac{n \sum xy - \sum x \sum y}{n \sum x^2 - (\sum x)^2}$			$a = \frac{\sum y - b \sum x}{n}$		
b=			a=		
b= 0.003585132			a= 0.1181291		
Y=0.118+0.0036X					
Error estándar del estimado		Varianza de la variable separable		Valor de correlación	
$Se = \sqrt{\frac{\sum y^2 - a \sum y - b \sum xy}{n - 2}}$		$S_y^2 = \frac{\sum y^2}{n} - \bar{y}^2$		$R^2 = 1 - \frac{Se^2}{S_y^2}$	
se=		1.8525		R^2= 0.8756155	

Figura 3. Valores de pendiente e intercepto. Fuente: Elaboración propia.

De esta forma se ha aplicado los modelos matemáticos en Excel de la regresión lineal simple por mínimos cuadrados para el caso de estudio y se ha encontrado el valor de correlación.

Para comprobar, utilizando el mismo Excel, se agrega un gráfico de dispersión y en la ventana emergente se selecciona agregar línea de tendencia, importante resaltar, considerando los datos, el comportamiento, indicativo, sobre la potencialidad, en cuanto a la conformación de los mismos como una línea recta y el valor alto de R² (Véase Figura 4):

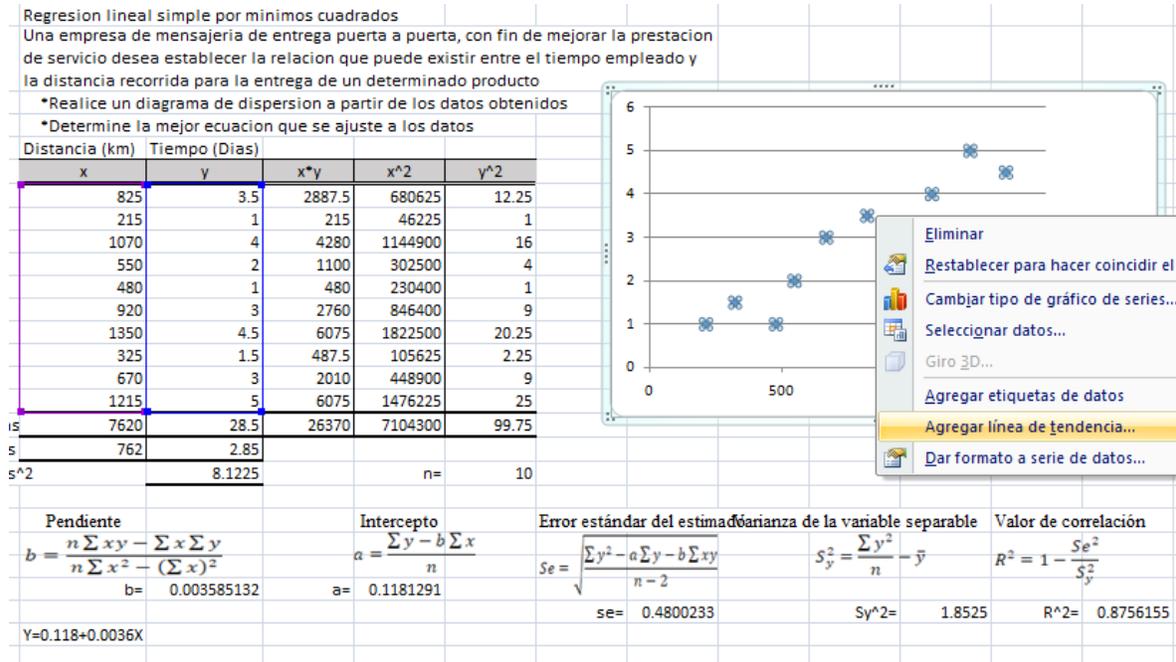


Figura 4. Gráfico de dispersión (Línea de tendencia). Fuente: Elaboración propia.

Se selecciona “Presentar ecuación en el gráfico” y “Presentar el valor de R cuadrado en el Gráfico” (Véase Figura 5).

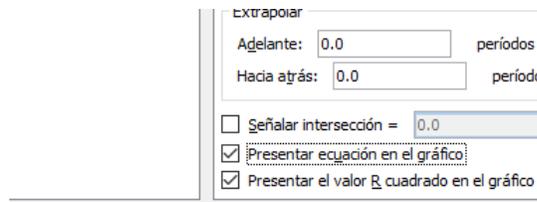


Figura 5. Ecuación y valor de R en el gráfico. Fuente: Elaboración propia.

Y se comprueban los resultados de aplicar las ecuaciones con la herramienta proporcionada por Excel (Véase Figura 6).

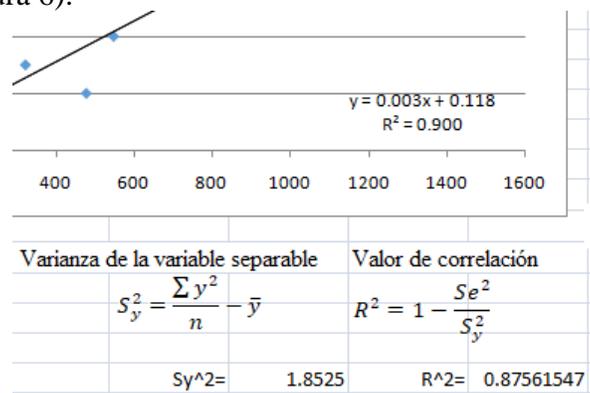


Figura 6. Comprobación de resultados mediante valor de R. Fuente: Elaboración propia.

Los resultados proporcionados en la gráfica de dispersión son similares a los que se obtuvieron aplicando los modelos matemáticos.

Excel es muy versátil, así se presenta un problema de regresión polinomial de orden 2. Partiendo del modelo propuesto de solución:

El modelo de la ecuación de regresión cuadrática

$$y = b_0 + b_1x + b_2x^2$$

Sistema de ecuaciones normales

$$\begin{aligned} b_0n + b_1 \sum x + b_2 \sum x^2 &= \sum y \\ b_0 \sum x + b_1 \sum x^2 + b_2 \sum x^3 &= \sum xy \\ b_0 \sum x^2 + b_1 \sum x^3 + b_2 \sum x^4 &= \sum x^2y \end{aligned}$$

El proceso en Excel (Véase Figura 7):

Regresion polinomial de orden 2						
Se llevó un experimento para determinar la distancia de frenado a diferentes velocidades de un nuevo modelo y se registraron los siguientes datos:						
Velocidad (Distancia de frenado (m)						
x	y	x*y	x^2	x^3	x^4	(x^2)y
30	15	450	900	27000	810000	13500
40	25	1000	1600	64000	2560000	40000
50	35	1750	2500	125000	6250000	87500
60	65	3900	3600	216000	12960000	234000
70	100	7000	4900	343000	24010000	490000
80	150	12000	6400	512000	40960000	960000
Sumatorias	330	390	26100	19900	1287000	87550000
n=	6					
Sistema de ecuaciones normales						
$b_0n + b_1 \sum x + b_2 \sum x^2 = \sum y$						
$b_0 \sum x + b_1 \sum x^2 + b_2 \sum x^3 = \sum xy$						
$b_0 \sum x^2 + b_1 \sum x^3 + b_2 \sum x^4 = \sum x^2y$						

Figura 7. Problema de regresión polinomial de orden 2. Fuente: Elaboración propia.

Se puede ver, debe resolver un sistema de 3 ecuaciones y 3 incógnitas, para ello Excel permite resolverlo por medio de la multiplicación de la inversa, esta resolución se logra por medio de los comandos de Excel MINVERSA (Véase Figura 8) y MMULT (Véase Figura 9).

Estas dos funciones, permiten reducir el tiempo de cálculo y generar la solución apropiada, agregando en forma, la obtención de otros estadísticos y parámetros de importancia, en el desarrollo de las pruebas de hipótesis, entre ellos, la eficacia del modelo, eficacia de las constantes, del coeficiente de correlación, con la determinación del estimador de la varianza poblacional (Véase Figura 9):

Sistema de ecuaciones normales						
$b_0 n$	$+ b_1 \sum x$	$+ b_2 \sum x^2$	$=$	$\sum y$		
$b_0 \sum x$	$+ b_1 \sum x^2$	$+ b_2 \sum x^3$	$=$	$\sum xy$		
$b_0 \sum x^2$	$+ b_1 \sum x^3$	$+ b_2 \sum x^4$	$=$	$\sum x^2 y$		
Sistema de ecuaciones resultantes						
6	330	19900	b0	=	390	
330	19900	1287000	b1	=	26100	
19900	1287000	87550000	b2	=	1825000	
b0	=	=MINVERSA(B26:D28)	0.007321429	390	65.28571	
b1	=	-0.83679	0.032982	-0.00029464	26100	= -3.23571
b2	=	0.007321	-0.00029	2.67857E-06	1825000	0.053571

Figura 8. Función MINVERSA. Fuente: Elaboración propia.

Sistema de ecuaciones resultantes						
6	330	19900	b0	=	390	
330	19900	1287000	b1	=	26100	
19900	1287000	87550000	b2	=	1825000	
b0	=	21.90714	-0.83679	0.007321429	390	= =MMULT(D30:F32,
b1	=	-0.83679	0.032982	-0.00029464	26100	G30:G32)
b2	=	0.007321	-0.00029	2.67857E-06	1825000	0.053571

Figura 9. Función MMULT. Fuente: Elaboración propia.

El modelo de la ecuación de regresión cuadrática

$$y = b_0 + b_1x + b_2x^2$$

Y con los coeficientes obtenidos en Excel se tiene (Véase Figura 10):

$$y = 65.2857143 - 3.23571429x + 0.05357143x^2$$

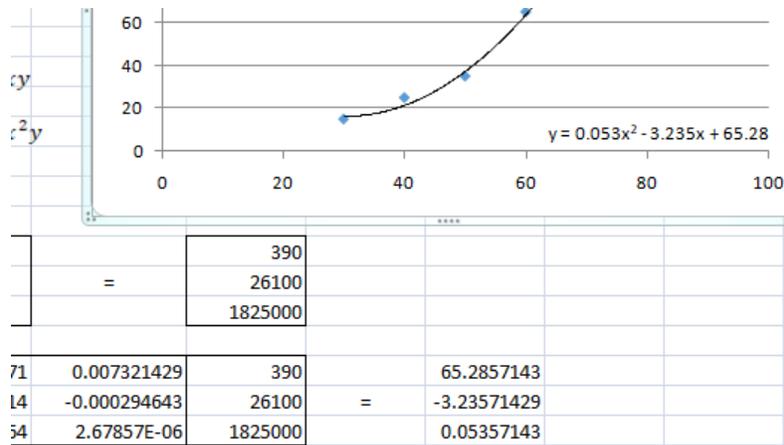


Figura 10. Coeficientes obtenidos. Fuente: Elaboración propia.

Con un procedimiento de añadir la gráfica de dispersión, pero en esta ocasión con una línea de tendencia polinómica de orden 2 y el modelo de la ecuación en la gráfica se puede corroborar el resultado.

En congruencia, el aprendiz, logra mayor dominio del conocimiento, habilidades y destreza con aptitudes relevantes muy superiores a aquellos que utilizan exclusivamente la enseñanza tradicional (Manual).

CONCLUSIONES

Las herramientas didácticas, propuestas para la formación en ciencias matemáticas y de estadística y, más en las instituciones superiores tecnológicas, es sugestivo que quien forma se asista preliminarmente, recurriendo en la simbolización gráfica y la expresión verbal y, afinar con presentaciones, figuras u otras herramientas, afirmadoras de la teoría, el procedimiento de enseñanza-aprendizaje, no termina en el aula, existirá la continuidad y acotamiento del grado de aprendizaje y, sí es necesario reforzar el conocimiento del aprendiz, exhortándole en el deseo de prevalecer, agregando en los procesos de indecisión o inhabilidad de progresar, presentarse a la asesoría.

La perspectiva de beneficio está en integrar la secuenciación y avance de la teoría, puntualizando apropiadamente la relación entre un nivel anterior y el resto de ellos, conceptualizando debidamente y dar respuesta al qué, el cómo y el para qué, del objeto de estudio de cada una de las diversas áreas de la ciencia y por ende de la filosofía, de este modo, el aprendiz retiene con mayor facilidad y dominio el conocimiento acerca del tema presentado. La derivación de las expresiones matemáticas, aún en Probabilidad y Estadística es arte y se transfiere su uso como conocimiento.

BIBLIOGRAFÍA

Cardona, J., Rojas, L. y Mesa, F. (2008). *Excel aplicaciones en álgebra, estadística, probabilidad y física*. Colombia: ECOE Ediciones.

Spiegel, M. y Stephens, L. (2011). *Estadística* (4ª ed.). México: McGraw-Hill, Interamericana

Tecnológico Nacional de México (2010). *Estadística inferencial 2. Programa de Estudios para la carrera de Ingeniería en Gestión Empresarial*. Recuperado de: <https://aguascalientes.tecnm.mx/reticulas-carreras/ige/Programas-IGEM-2009-201/Estadistica-Inferencial-II-IGE-2009.pdf>

Tecnológico Nacional de México (2016). *Estadística inferencial 1. Programa de Estudios para la carrera de Ingeniería Industrial*. Recuperado de: <http://www.apizaco.tecnm.mx/Oficial/IND/pdf/AEF1024.pdf>