

Aplicación de modelos predictivos y correlaciones

Application of predictive models and correlations

Luis E. Monreal-Pitones ^a

Abstract:

The use of predictive models corresponds to a mathematical practice whose use is applicable to both social and exact sciences. Working with predictions helps any organization to know the results that are expected to be achieved in a certain period, through a previous history of recorded events. Today there are computer algorithms that process information in an automated way and allow calculations to be processed to obtain results that estimate a future scenario. The methods used in this report are: Moving Average, Absolute Variation, Linear Regression and Exponential Smoothing.

Keywords:

Linear regression, Forecasts, Trends, Correlation

Resumen:

El uso de modelos predictivos corresponde a una práctica matemática cuyo uso es aplicable a ciencias tanto sociales como exactas. Trabajar con predicciones ayuda a toda organización a conocer los resultados que se esperan alcanzar en un determinado periodo, mediante un historial previo de sucesos registrados. Hoy en día existen algoritmos informáticos que procesan la información en forma automatizada y permiten procesar los cálculos para la obtención de resultados que estimen un futuro escenario. Los métodos que se utilizan en este reporte son: Promedio móvil, Variación Absoluta, Regresión Lineal y Suavización Exponencial.

Palabras Clave:

Regresión lineal, Pronósticos, Tendencias, Correlación

Introducción

El presente proyecto muestra diversos modelos matemáticos utilizados para el uso de estimaciones mediante el análisis de tendencias y comportamientos de una serie de datos históricos. Los métodos presentados en este proyecto son Promedio Móvil, Variación Absoluta, Regresión Lineal, Suavización Exponencial y Correlación de Pearson. Cada demostración se encuentra ejemplificada mediante un caso relacionado con la actividad profesional relacionada hacia la práctica comercial.

Uso de promedios móviles

Un pronóstico por promedio móvil permite estimar una cantidad a partir de una serie de datos consecutivos de periodos anteriores. Los promedios móviles por lo regular se realizan a 3 y 6 meses y su función se enfoca a:

- Proyectar cifras esperadas a un periodo dado
- Comparar la relación que existe en un periodo y otro
- Establecer cantidades mínimas a alcanzar en determinada acción

$$\text{Promedio móvil} = \frac{\sum x}{n}$$

La x representa cada uno de los valores previos al periodo que se desea estimar. La n representa el total de valores previos al periodo se desea estimar. En ambos casos la elección de valores puede ser 3 o 6 dependiendo de la proyección deseada, número total de valores o escenario en curso.

^a Autor de Correspondencia, Universidad Tecnológica Santa Catarina, <https://orcid.org/0000-0002-4635-1880>, Email: luise.monreal@utsc.edu.mx

Promedio móvil a 3 meses

Ejemplo de una proyección de unidades a vender para el mes de octubre.

La siguiente tabla muestra el historial de ventas de unidades pertenecientes al año 2020. En ella podemos observar un comportamiento que muestra tanto incrementos como descensos en los 9 meses recopilados. El objetivo es utilizar la función de Promedio móvil para estimar las unidades a vender durante el mes de octubre.

Año 2020	
Mes	Unidades vendidas
enero	3,000
febrero	3,600
marzo	4,000
abril	3,900
mayo	5,900
junio	3,200
julio	4,000
agosto	7,600
septiembre	8,100
octubre	

Tabla 1. Ventas del periodo enero-septiembre 2020

Si realizamos un Promedio móvil utilizando los últimos 3 meses, el resultado sería el siguiente:

$$\text{Promedio móvil} = \frac{\sum x}{n} = \frac{(\text{julio} + \text{agosto} + \text{septiembre})}{3}$$

$$\text{Promedio móvil} = \frac{\sum x}{n} = \frac{(4,000 + 7,600 + 8,100)}{3}$$

$$\text{Promedio móvil} = \frac{\sum x}{n} = \frac{(19,700)}{3}$$

$$\text{Promedio móvil} = \frac{\sum x}{n} = \frac{(19,700)}{3}$$

$$\text{Promedio móvil} = 6,566.66$$

El resultado obtenido pasa a redondearse para tener la cantidad de 6,567 siendo esta la cifra esperada de unidades vendidas para el mes de octubre.

julio	4,000
agosto	7,600
septiembre	8,100
octubre	6,567

Tabla 2. Proyección del mes de octubre

Promedio móvil a 6 meses

Si realizamos un Promedio móvil utilizando los últimos 6 meses, el resultado sería el siguiente:

$$\text{Promedio móvil} = \frac{\sum x}{n} = \frac{(\text{abril} + \text{mayo} + \text{junio} + \text{julio} + \text{agosto} + \text{septiembre})}{6}$$

$$\text{Promedio móvil} = \frac{\sum x}{n} = \frac{(3,900 + 5,900 + 3,200 + 4,000 + 7,600 + 8,100)}{6}$$

$$\text{Promedio móvil} = \frac{\sum x}{n} = \frac{(32,700)}{6}$$

$$\text{Promedio móvil} = 5,450$$

El resultado obtenido corresponde a 5,450 siendo esta la cifra esperada de unidades vendidas para el mes de octubre, tomando como referencia los últimos 6 periodos (enero-septiembre).

Abril	3,900
Mayo	5,900
Junio	3,200
Julio	4,000
Agosto	7,600
Septiembre	8,100
Octubre	5,450

Tabla 3. Proyección del mes de octubre con Promedio Móvil a 6 meses

En ambos casos (3 y 6 meses) el resultado proyectado siempre será inferior al último periodo tomado, razón que justifica que el uso de Promedios móviles no debe tomarse como una cifra máxima para alcanzar, sino como un resultado mínimo esperado o bien como un comparativo de la cifra esperada contra la cifra realizada.

Año 2020			
Mes	Unidades vendidas	Pronóstico con Promedio móvil a 3 meses	Pronóstico con Promedio móvil a 6 meses
enero	3,000		
febrero	3,600		
marzo	4,000		
abril	3,900		
mayo	5,900		
junio	3,200		
julio	4,000		
agosto	7,600		
septiembre	8,100		
octubre		6,567	5,450

Tabla 4. Comparación de pronósticos con periodos de 3 y 6 meses

Promedio móvil para comparar

Este método permite obtener un promedio del comportamiento que tuvieron las cifras periodo tras periodo en un año determinado, para de esta forma poder comparar los resultados obtenidos en un año próximo contra los resultados proyectados.

En el siguiente ejemplo se realizará un pronóstico de promedio móvil de 3 meses a partir del mes de abril hasta el mes de septiembre y después se comparará dicho resultado con las ventas obtenidas en el 2021.

Año 2020		Año 2021
Mes	Unidades vendidas	Unidades proyectadas
enero	3,000	
febrero	3,600	
marzo	4,000	
abril	3,900	3,533
mayo	5,900	3,833
junio	3,200	4,600
julio	4,000	4,333
agosto	7,600	4,367
septiembre	8,100	4,933

Tabla 5. Comparativo de ventas reales vs ventas proyectadas

Para obtener el pronóstico del mes de abril del 2021 se tomaron los meses enero, febrero y marzo del 2020 y así sucesivamente con el resto de los meses.

Luego de hacer el pronóstico por Promedio Móvil de cada mes se incluye en la tabla las unidades que realmente se vendieron durante el año 2021.

Año 2020		Año 2021	Año 2021
Mes	Unidades vendidas	Unidades proyectadas	Unidades vendidas
enero	3,000		2,500
febrero	3,600		1,900
marzo	4,000		3,000
abril	3,900	3,533	1,900
mayo	5,900	3,833	2,100
junio	3,200	4,600	5,000
julio	4,000	4,333	4,900
agosto	7,600	4,367	3,900

Tabla 6. Comparativo de unidades vendidas vs unidades proyectadas

Como puede apreciarse en la tabla, en algunos periodos hubo una diferencia negativa entre las ventas proyectadas y las ventas reales. El promedio móvil ayudó a determinar si el comportamiento que tuvieron las ventas durante el año 2021 era el pronosticado tomando como referencia el comportamiento que se tuvo en el transcurso del año 2020.

Promedio de Variación Absoluta

Este método es requerido cuando se desea pronosticar las ventas a corto, mediano o largo plazo teniendo como base un mínimo de datos históricos. Su función consiste en sumar las variaciones que cada periodo tuvo para luego efectuar una división de esta suma entre el número de variaciones que se tuvieron. El resultado de esta operación será sumado a las ventas del último periodo registrado con el fin de usarse como venta pronosticada. En el siguiente ejemplo se presenta un Pronóstico de Ventas que utiliza el Promedio de Variación Absoluta.

Una compañía distribuidora de hardware ha registrado las siguientes ventas en su primer cuatrimestre del año.

Mes	Ventas Realizadas
Enero	\$ 45,000
Febrero	\$ 48,000
Marzo	\$ 39,500
Abril	\$ 43,900
Mayo	

Tabla 7. Ventas del periodo enero-abril

El objetivo es pronosticar las ventas del mes de Mayo y el año restante utilizando Promedios de Variación Absoluta aplicando la siguiente formula.

$$PVA = \frac{\text{Suma de variaciones}}{\text{Número de periodos}-1}$$

	Ventas Realizadas		Variaciones
Enero	\$ 45,000	De Enero a Febrero las ventas tuvieron una variación positiva de \$3,000	+ \$3000
Febrero	\$ 48,000		
Marzo	\$ 39,500	De Febrero a Marzo las ventas tuvieron una variación negativa de \$8,500	- \$8,500
Abril	\$ 47,900	De Marzo a Abril las ventas tuvieron	+ \$8,400

		una variación positiva de \$8,400	
			Suma de variaciones \$ 2,900

Tabla 8. Obtención de variaciones entre cada mes

Una vez obtenida la variación de cada periodo se procede a realizar la suma, la cual tiene como resultado la cantidad de \$2,900, para posteriormente sustituir ese valor en la fórmula del Promedio de Variación Absoluta.

$$PVA = \frac{\text{Suma de variaciones}}{\text{Número de periodos}-1}$$

$$PVA = \frac{2,900}{4-1} = \frac{2,900}{3} = 966.66$$

El resultado obtenido pasará a ser redondeado a 967 y será sumado al último periodo registrado del año con el fin de estimar el total de ventas anuales sumando dicha cantidad de manera consecutiva.

Al mes de Mayo se le sumará la cantidad de 967 que corresponde a la variación estadística que se tiene en cada periodo.

	Ventas Realizadas
Enero	\$ 45,000
Febrero	\$ 48,000
Marzo	\$ 39, 500
Abril	\$ 43, 900
Mayo	\$ 44,867
Junio	\$ 45,834
Julio	\$ 46, 801
Agosto	\$ 47,768
Septiembre	\$ 48, 735
Octubre	\$ 49, 702
Noviembre	\$ 50, 669
Diciembre	\$ 51, 636

Tabla 9. Proyección final con el Promedio de Variación Absoluta

Al resultado pronosticado (Mes de Mayo) de igual manera se le sumará el Promedio de Variación Absoluta para así poder pronosticar el resto del año basado en el comportamiento que tuvieron las ventas durante los primeros meses.

De igual manera se procede a estimar el total de ventas anuales sumando la cantidad vendida en los primeros meses más las cantidades pronosticadas para el resto de los años, lo cual da un resultado de \$231, 469.

Regresión Lineal

El uso de pronósticos por medio de la regresión lineal ayudará a las empresas en estimar las ventas para un periodo consecutivo cuando su tendencia sea creciente o decreciente. Su función se basa en determinar la relación que existe entre una variable y las series de tiempo que se tiene.

La fórmula para realizar un pronóstico de ventas utilizando el método de Regresión Lineal es la siguiente:

$$b = \frac{n \sum xt - [\sum x \sum t]}{n \sum t^2 - \sum [t]^2} \quad a = \bar{x} - b(\bar{t})$$

$$\hat{x} = a + b(\hat{t})$$

Donde **n** representa el número de periodos que se tienen, la letra **t** indica el periodo y la letra **x** corresponde a las ventas de cada periodo

En el siguiente ejemplo se presenta un Pronóstico de Ventas que utiliza el método de Regresión Lineal.

Una agencia de viajes desea saber las ventas estimadas para el séptimo mes (julio). Determina el pronóstico de Ventas usando el método de Regresión Lineal.

	Mes	Viajes vendidos
Enero	1	20
Febrero	2	25
Marzo	3	30
Abril	4	25
Mayo	5	40
Junio	6	39

Tabla 10. Ventas del periodo enero-junio

Se obtienen las sumatorias y los cuadrados que se usarán en la formula. Cada mes tendrá el valor de tiempo consecutivo y será representado con la letra t (1,2,3,4,5,6), donde el objetivo es pronosticar el periodo 7. Los viajes vendidos serán representados por la letra x.

$\sum t$	$\sum x$	$\sum xt$	$\sum t^2$	$\sum [t]^2$
Mes	Viajes vendidos			
1	20	20	1	(21)(21)
2	25	50	4	
3	30	90	9	
4	25	100	16	
5	40	200	25	
6	39	234	36	
21.00	179.00	694.00	91.00	441.00

Tabla 11. Obtención de sumatorias

Para obtener $\sum xt$ se multiplica cada valor de t por cada valor de x y se realiza la suma del resultado de cada uno de estos.

Para obtener $\sum t^2$ se suma cada valor de t elevado al cuadrado

Para obtener $\sum [t]^2$ se eleva la suma todos los valores de t , al cuadrado.

Sustituir los valores encontrados en la primera fórmula para encontrar el valor de b , donde n representa el número de periodos actuales, es decir: El número de meses (6) que se tienen como datos históricos.

$$b = \frac{n \sum xt - [\sum x - \sum t]}{n \sum t^2 - \sum [t]^2}$$

$$b = \frac{6(694) - [(179)(21)]}{6(91) - 441}$$

$$b = \frac{4164 - 3759}{546 - 441}$$

$$b = \frac{405}{105} \quad b = 3.85$$

Se encuentra el valor de a . El símbolo \bar{x} indica el promedio de la sumatoria de las x . El símbolo \bar{t} indica el promedio de la sumatoria de las t .

$$\bar{x} = \frac{\sum x}{n} = \frac{179}{6} = 29.83$$

$$\bar{t} = \frac{\sum t}{n} = \frac{21}{6} = 3.50$$

$$a = \bar{x} - b(\bar{t})$$

$$a = 29.83 - [(3.85)(3.50)]$$

$$a = 29.83 - 13.50$$

$$a = 16.33$$

Una vez que se encontró a , se pronostica el resultado donde \hat{t} representa el periodo a encontrar. En este caso el periodo a encontrar es el número 7, ya que se tenían 6 periodos anteriores que corresponden a los meses de enero-junio, dando como resultado la proyección de 43 viajes a vender durante el mes de julio.

$$\hat{x} = a + b(\hat{t})$$

$$\hat{x} = 16.33 + 3.85(7)$$

$$\hat{x} = 16.33 + 27$$

$$\hat{x} = 43.33$$

Mes	Viajes vendidos	Viajes proyectados
1	20	
2	25	
3	30	
4	25	
5	40	
6	39	
7		43

Tabla 12. Proyección del periodo 7

Suavización exponencial

También conocida como alisamiento exponencial busca ajustar el pronóstico mediante una corrección obtenida tras la venta real que será comparada con la venta que se pronosticó con anterioridad. Es utilizado por muchas empresas como un avance del promedio móvil.

Para obtener la nueva estimación se requiere de tres datos: El último periodo pronosticado, la cantidad pronosticada y el coeficiente de suavización.

La fórmula que representa el modelo de Alisamiento Exponencial es la siguiente:

$$\hat{x}_t = \hat{x}_{t-1} + (\alpha(x_{t-1} - \hat{x}_{t-1}))$$

Donde

\hat{x}_{t-1} Representa el pronóstico de ventas del periodo anterior

x_{t-1} Representa las ventas reales del periodo anterior

α es la constante de suavización. Debe tener un valor de entre 0,0 y 1,0

n representa el número de periodos que se tienen en el pronóstico

Calcular el coeficiente de suavización

El coeficiente de suavización se asigna su valor dependiendo la estabilidad que haya tenido la demanda

a través de año, es decir su las ventas no han tenido un crecimiento exponencial en su comportamiento, se le asigna un coeficiente de suavización pequeño (0,10 por ejemplo), no obstante si la compañía presenta cambios elevados en la demanda se recomienda usar constantes de suavización un poco más elevadas con el fin otorgarle más prioridad a la demanda reciente y estimar un resultado más preciso. Veamos el siguiente ejemplo. Una compañía de seguros pronostico para su primer periodo una demanda de 950 seguros contratados, sin embargo, la demanda real fue de 1,600. Calcular el pronóstico para el segundo periodo utilizando una constante de suavización (0,40).

$$\hat{x}_t = \hat{x}_{t-1} + (\alpha(x_{t-1} - \hat{x}_{t-1}))$$

$$\hat{x}_t = 950 + (0.40(1600 - 950))$$

$$\hat{x}_t = 950 + (0.40(650))$$

$$\hat{x}_t = 950 + 260$$

$$\hat{x}_t = 1210$$

Al tratarse de una demanda que si tuvo un incremento considerable con respecto a lo pronosticado se optó por utilizar un coeficiente de suavización de (0,40). Una vez realizado el ejercicio, se concluye el análisis con la siguiente tabla:

Periodo	Venta Real	Venta Pronosticada
Enero	1,600	950
Febrero		1,210

Tabla 13. Nuevo pronóstico con Suavización Exponencial

Correlaciones

En aquellas prácticas que conlleven la toma de decisiones basadas en un historial de hechos, el uso de Correlaciones ayuda en que se pueda medir el impacto que tuvo una serie de acciones con los resultados generados de dichas acciones.

Existen dos formas de conocer si existe una correlación entre dos grupos de variables: La primera es mediante el análisis de los datos en una Gráfica de Dispersión y la segunda forma es realizando el cálculo del Coeficiente de Correlación de Pearson. Ambos métodos son utilizados en conjunto como factor de comprobación.

Correlación por gráfica de dispersión

La correlación entre datos se calcula graficando los puntos correspondientes en los ejes X y Y, con el fin de

analizar si los datos X incrementan a medida que aumenta la frecuencia de los datos Y.

El siguiente ejemplo muestra un recuento de la inversión en Redes Sociales que una compañía hizo durante 10 semanas y al mismo tiempo los ingresos obtenidos. El objetivo es saber si existe una relación entre la inversión realizada y el aumento de las ventas.

Hipótesis inicial:

A mayor presupuesto de inversión, mayor son las ventas.

	X	Y
Semana	Inversión en publicidad (\$)	Ingresos obtenidos (\$)
1	3,500	150,000
2	3,600	158,000
3	4,000	170,000
4	2,000	90,000
5	1,000	75,000
6	5,000	173,000
7	5,500	170,000
8	5,000	185,000
9	5,000	180,000
10	4,000	165,000

Tabla 14. Inversión en publicidad e ingresos (10 semanas)

El primer paso consiste en graficar en un plano cartesiano los valores de la tabla anterior a manera de coordenadas, siendo la Inversión en publicidad las coordenadas Y y los ingresos las coordenadas X.

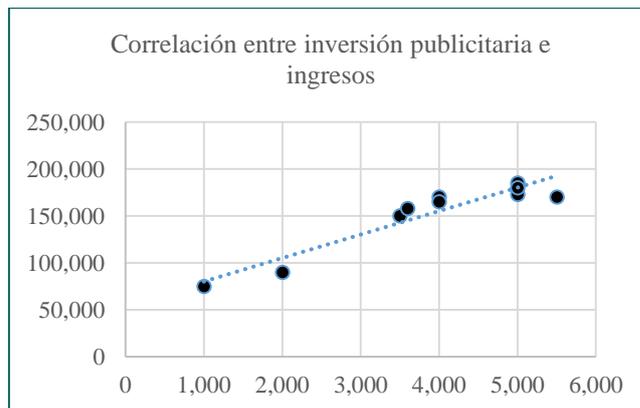


Tabla 15. Correlación entre Inversión en publicidad e Ingresos

Analizando la gráfica obtenida podemos notar que a medida que la inversión en publicidad aumenta también aumentan los ingresos generados, obteniendo una Correlación Positiva.

Si al Gráfico de Dispersión le trazamos una línea de tendencia, podemos visualizar con mayor precisión la

relación positiva que existe entre ambas variables. La inclinación de la línea de tendencia va a depender del comportamiento que tengan los datos.

Coefficiente de Correlación de Pearson

El Coeficiente de Correlación de Pearson es una medida de recuento cuyo proceso permite determinar la relación o intensidad que existe entre dos variables, de esta forma poder conocer si el comportamiento de un conjunto de datos influye sobre otro, ambos grupos de datos pertenecientes al mismo escenario.

La fórmula para calcular el Coeficiente de Correlación de Pearson es la siguiente:

$$\text{Coeficiente de Correlación} = \frac{\sum xy}{\sqrt{(\sum x^2)(\sum y^2)}}$$

Se utilizará el mismo ejemplo de la correlación gráfica (punto anterior) para su comprobación con la fórmula del Coeficiente de Correlación de Pearson.

Semana	X Inversión en publicidad (\$)	Y Ingresos obtenidos (\$)
1	3,500	150,000
2	3,600	158,000
3	4,000	170,000
4	2,000	90,000
5	1,000	75,000
6	5,000	173,000
7	5,500	170,000
8	5,000	185,000
9	5,000	180,000
10	4,000	165,000

Tabla 16. Inversión en publicidad e ingresos (10 semanas)

Lo primero es diferenciar las variables minúsculas y mayúsculas (X, x, Y, y), para ello se aplican las siguientes formulas:

$$x = X - \bar{X}$$

$$y = Y - \bar{Y}$$

El promedio de las X corresponde a $\bar{X} = \frac{38600}{10} = 3860$

El promedio de las Y corresponde a $\bar{Y} = \frac{1516000}{10} = 151600$

Enseguida se resta a cada valor de X su respectivo promedio, para obtener x y posteriormente la sumatoria de estas. El mismo proceso se repetirá con la letra Y. Una forma más dinámica de realizar estas operaciones es mediante la siguiente tabla.

	X	Y	x	y			
Se ma na	Inv ersi ón	Ingr eso s	$x = X - \bar{X}$	$y = Y - \bar{Y}$	xy	x^2	y^2

1	3,500	150,000	-360	-1,600	57600	129600	256000
2	3,600	158,000	-260	-640	166400	67600	409600
3	4,000	170,000	140	1840	257600	19600	338560
4	2,000	90,000	-1860	-6160	114560	34560	3794560
5	1,000	75,000	-2860	-7660	21900	81760	5867560
6	5,000	173,000	1140	2140	243960	12960	457960
7	5,500	170,000	1640	1840	301760	26860	338560
8	5,000	185,000	1140	3340	380760	12960	1115560
9	5,000	180,000	1140	2840	323760	12960	806560
10	4,000	165,000	140	1340	187600	19600	179560
Totales					4620000	184600	12942000

Teniendo los totales para cada letra, se remplazan los valores en la fórmula del Coeficiente de Correlación de Pearson.

$$\text{Coeficiente de Correlación} = \frac{462040000}{\sqrt{(18464000)(12942400000)}}$$

$$\text{Coeficiente de Correlación} = \frac{462040000}{\sqrt{238968473600000000}}$$

$$\text{Coeficiente de Correlación} = 0.94$$

El resultado del Coeficiente de Correlación es 0.94, lo cual indica que, de acuerdo con la tabla de valores de dicho coeficiente, existe una correlación muy alta entre los ingresos y la inversión en publicidad. Esta correlación pudo demostrarse en el punto anterior donde visualmente la tendencia de datos en dispersión iba en forma ascendente.

Valor obtenido	Significado
-1	Correlación negativa grande y perfecta
-0.9 a -0.99	Correlación negativa muy alta

-0.7 a -0.89	Correlación negativa alta
-0.4 a -0.69	Correlación negativa moderada
-0.2 a -0.39	Correlación negativa baja
-0.01 a -0.19	Correlación negativa muy baja
0	No hay correlación
0.01 a 0.19	Correlación positiva muy baja
0.2 a 0.39	Correlación positiva baja
0.4 a 0.69	Correlación positiva moderada
0.7 a 0.89	Correlación positiva alta
0.9 a 0.99	Correlación positiva muy alta
1	Correlación positiva grande y perfecta

Referencias

- [1] Hanke, J. (2010). Pronósticos en los negocios. Pearson.
- [2] Anderson, D. R., Sweeney, D., Williams, T. A., Camm, J. D., & Martín, K. (2011). Métodos cuantitativos para los negocios 11va Edición. Ciudad de México: Cengage Learning.
- [3] Lujan, D. (2009). El qué y cómo de los pronósticos en las ventas. Gasca.
- [2] Betancourt, D. (14 de Septiembre de 2021). ¿Cómo usar la suavización exponencial simple como pronóstico de producción? Recuperado de Ingenio Empresa: <https://www.ingenioempresa.com/suavizacion-exponencial-simple/>
- [4] Montemayor, E. (2015). Métodos de pronósticos para los negocios. Monterrey: Tecnológico de Monterrey.
- [5] Minitab. (29 de marzo de 2022). Interpretar los resultados clave para una correlación. Obtenido de Soporte de MiniTab: <https://support.minitab.com/es-mx/minitab/18/help-and-how-to/statistics/basic->