

La Correlación de Pearson y el proceso de regresión por el Método de Mínimos Cuadrados

Germán Fiallos

gfiallos2@indoamerica.edu.ec

Facultad de Jurisprudencia, Ciencias Políticas y Económicas

Carrera de Derecho

Universidad Tecnológica Indoamérica

Ambato – Ecuador

RESUMEN

Se realiza un breve análisis de las características de los gráficos de dispersión y la relación que existe entre las dos variables utilizando el proceso de análisis de correlación r de Pearson, sus características directas o indirectas y la verificación de su resultado con ayuda del Excel, posteriormente se utiliza el proceso de regresión por mínimos cuadrados que nos permite determinar la ecuación lineal de mejor ajuste que identifique a todos los datos de la información representada ente las variables del gráfico de dispersión, modelo matemático que nos permite inferir resultados del comportamiento entre las variables, de igual forma se realiza una verificación del modelo matemático de la ecuación lineal con ayuda del Excel.

Palabras clave: Correlación; Dispersión; Regresión

Pearson's Correlation and the regression process by the Least Squares Method

ABSTRACT

A brief analysis of the characteristics of the scatter charts and the relationship between the two variables is carried out using the Pearson r correlation analysis process, its direct or indirect characteristics and the verification of its result with the help of Excel, subsequently The least squares regression process is used that allows us to determine the linear equation of best fit that identifies all the data of the information represented between the variables of the scatter graph, a mathematical model that allows us to infer results of the behavior between the variables, in the same way, a verification of the mathematical model of the linear equation is carried out with the help of Excel.

Keywords: Correlation; Dispersion; Regression

Artículo recibido: 05 de Abril 2021

Aceptado para publicación: 28 de Mayo 2021

Correspondencia gfiellos2@indoamerica.edu.ec

Conflictos de Interés: Ninguna que declarar

1. INTRODUCCIÓN

El estudio de la función lineal, especialmente sus características como pendiente, gráfico y modelo matemático tiene su aplicación o evolución en el estudio de correlaciones y regresiones lineales, utilizando herramientas como Excel o procesos analíticos como el método de los mínimos cuadrados, podemos determinar las características de la relación que existe entre ellas; establecer un nexo entre estos dos temas que se estudian en matemática como son funciones lineales y en Estadística las correlaciones y regresiones por mínimos cuadrados, permitirá justificar la interdisciplinariedad en el estudio de estos capítulos que de no ser relacionados, se presentarían como totalmente independientes entre sí pero como vemos tienen un mismo objetivo. El obtener información del tipo de relación que existe entre dos variables utilizando las herramientas de la matemática como las funciones lineales o en estadística las correlaciones r de Pearson y el método de regresiones por el proceso de los mínimos cuadrados, con ayuda de Excel, nos permite verificar procesos analíticos, resultados y determinar los modelos matemáticos que nos permiten obtener conclusiones o inferir resultados con respaldo científico.

1.1. Propósito

Compartir procesos matemáticos sobre el análisis de correlación entre 2 variables, realizar el análisis de regresión por procesos de mínimos cuadrados y verificar sus respuestas con ayuda de Excel, establecer las características de la función lineal aplicadas al estudio de correlaciones y regresiones lineales.

1.2. Resultado del aprendizaje

Aplica con probidad los algoritmos correspondientes para realizar el gráfico de dispersión y establecer el nivel de correlación entre variables mediante la utilización de coeficiente de r de Pearson; deduce la ecuación de regresión correspondiente utilizando el proceso de los mínimos cuadrados lo cual le permite inferir resultados, los cuales son verificados con ayuda de Excel.

1.3. Diagrama de aprendizaje

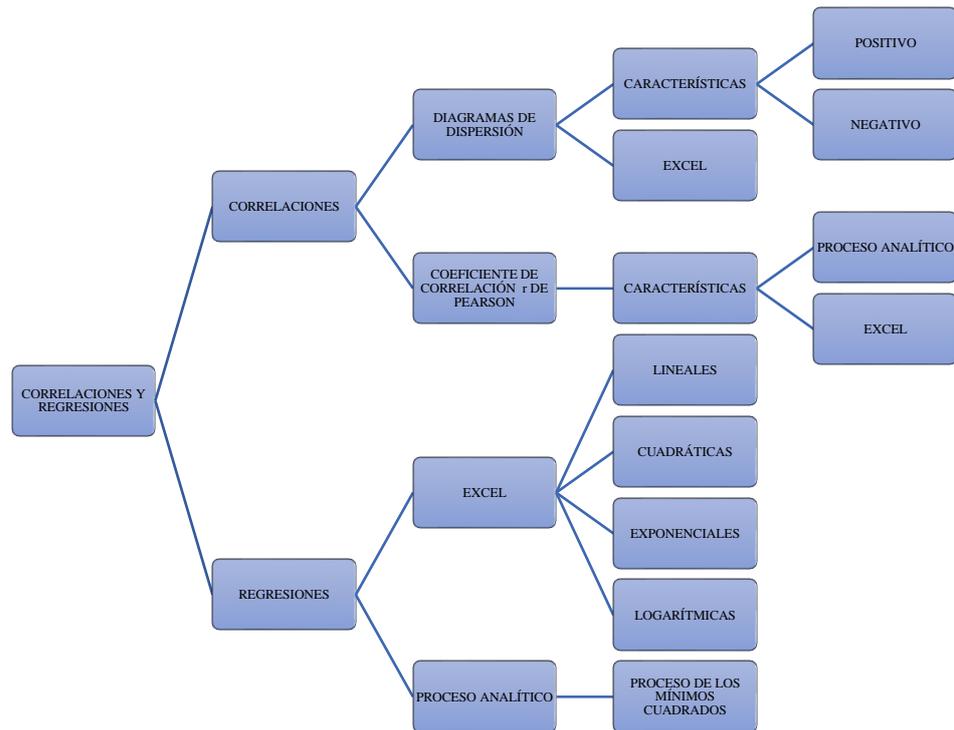


Gráfico elaborado por Germán Fiallos

2. ESTRATEGIAS METODOLÓGICAS O MATERIALES Y MÉTODOS

2.1. Diagrama de dispersión de puntos

- **Diagrama de dispersión** o gráfico de **dispersión** es un tipo de **diagrama** matemático que utiliza un sistema de referencia bidimensional, para mostrar los valores y la relación que existe entre dos variables de un conjunto de datos obtenidos en una investigación.

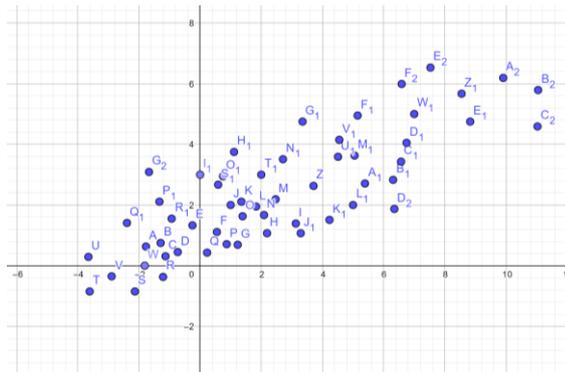
2.2. Características de los diagramas de dispersión de puntos

- Tiene dos variables cuantitativas, siendo (x) la variable llamada independiente y (y) la variable dependiente
- Los puntos no se unen entre si
- No es tabla de frecuencias

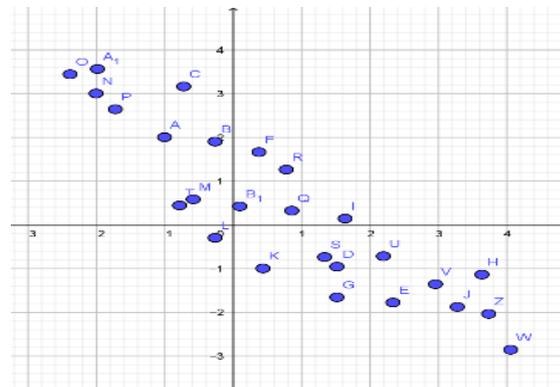
2.3. Tipos de diagramas de dispersión de puntos

El modelo de los datos es indicativo del tipo de relación entre las dos variables:

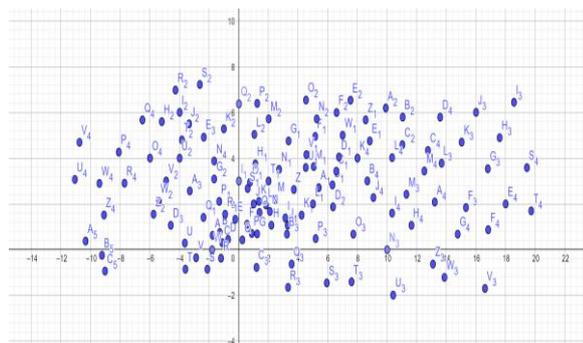
- Relación positiva entre las variables



- Relación negativa entre las variables



- No existe relación entre las variables



2.4. Coeficiente de correlación r de Pearson

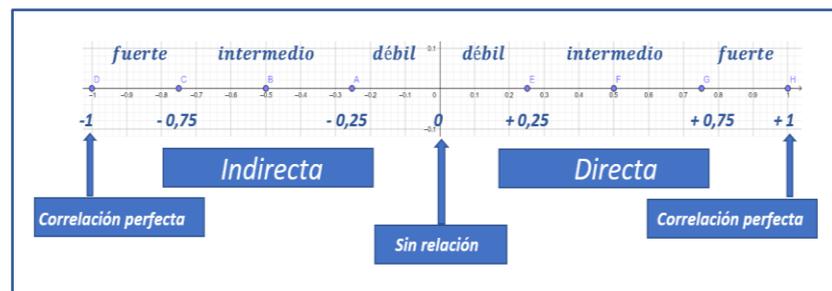
El índice numérico más común usado para medir una correlación es el “coeficiente de Pearson”. El coeficiente de Pearson (también llamado coeficiente de correlación del producto-momento), se representa con el símbolo r y proporciona una medida numérica de la correlación entre dos variables cuantitativas.

2.4.1. Características principales

También llamado *correlación r Pearson*, mide la naturaleza y fuerza entre dos variables cuantitativas, que permite describir la relación entre dos variables (correlación):

▪ **Características del coeficiente r de Pearson:**

- 1) Nos indica si dos variables están correlacionadas o no, $y = f(x)$
- 2) El *coeficiente r de Pearson* indica la fuerza de la aparente relación
- 3) El *coeficiente r de Pearson* nos indica si la aparente relación es positiva o negativa.
- 4) El signo del *coeficiente r de Pearson* nos indica la naturaleza de la correlación entre las variables
- 5) El valor del *coeficiente r de Pearson* denota la fuerza o intensidad la correlación entre las variables
- 6) Si el signo de la correlación es positivo, significa que la relación es directa (Función creciente donde un incremento en una variable está asociado con el incremento de la otra variable; una disminución de una variable está asociado con la disminución de la otra variable).
- 7) Si el signo de la correlación es negativo, significa una relación **inversa** o indirecta (Función decreciente significando que el incremento en una variable está asociado con una disminución de la otra variable).
- 8) El valor del *coeficiente r Pearson* está entre (-1) y (+1)
- 9) El valor del *coeficiente r Pearson* denota la fuerza de la asociación como se ilustra en el siguiente diagrama.



Elaborado por Germán Fiallos

1. Si $r = 0$ significa que no hay asociación o correlación entre las dos variables.
2. Si $0 < r < 0.25$ = débil correlación.
3. Si $0.25 \leq r < 0.75$ = intermedia correlación.
4. Si $0.75 \leq r < 1$ = fuerte correlación.
5. Si $r = \pm 1$ = perfecta correlación.
6. ¿Cómo calcular el coeficiente de correlación simple r Pearson?

- ✓ Utilizamos la ecuación de correlación lineal de Pearson

$$r = \frac{\sum xy - \frac{\sum x \sum y}{n}}{\sqrt{\left(\sum x^2 - \frac{(\sum x)^2}{n}\right) \left(\sum y^2 - \frac{(\sum y)^2}{n}\right)}}$$

EJEMPLO 1:

Se realiza una investigación para establecer la relación que existe entre la edad de unos niños y su peso; se selecciona una muestra de 6 individuos, los datos de su edad en años y peso en kilogramos fue registrada como se muestra en la siguiente tabla. Se requiere elaborar un diagrama de dispersión y encontrar la correlación entre edad y peso. Utilice el coeficiente de correlación *r* *Pearson's*

	<i>Edad</i> (años)	<i>Peso</i> (kg)
1	7	12
2	6	8
3	8	12
4	5	10
5	6	11
6	9	13

Tabla ejemplo 1

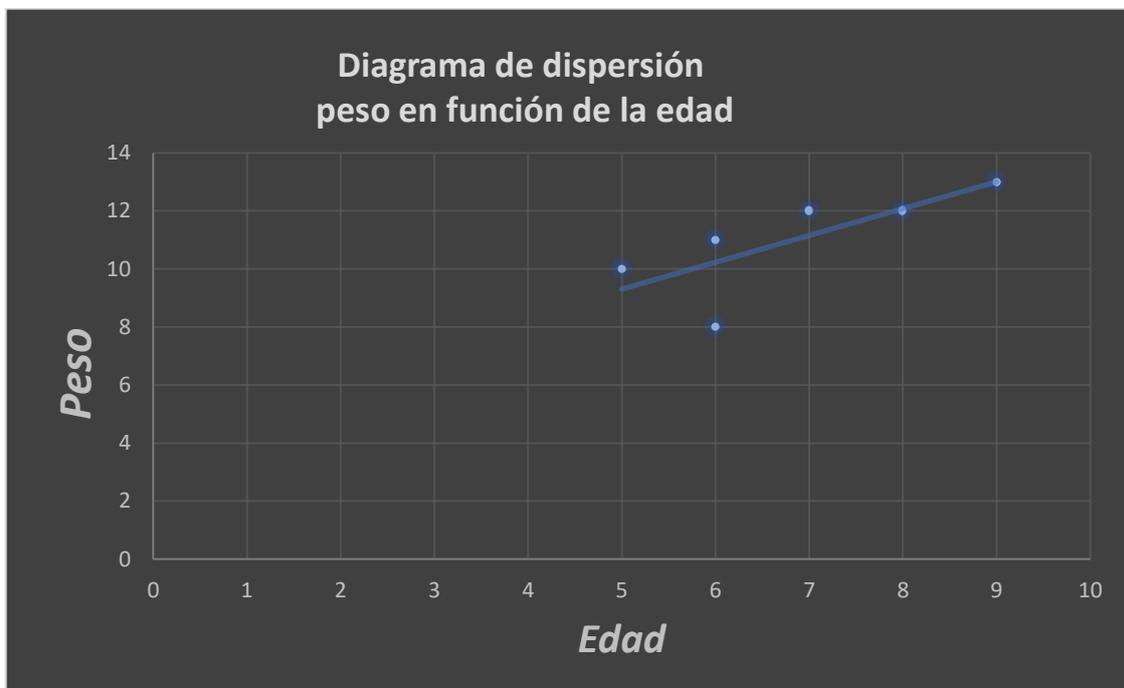


Gráfico elaborado por Germán Fiallos

Las dos variables son de tipo cuantitativo, una variable (edad) es llamada independiente; la otra (peso) es llamada dependiente con notación de variable y, para encontrar la relación entre edad y peso, calculamos el coeficiente de correlación simple, usando la siguiente ecuación:

$$r = \frac{\sum xy - \frac{\sum x \sum y}{n}}{\sqrt{\left(\sum x^2 - \frac{(\sum x)^2}{n}\right) \left(\sum y^2 - \frac{(\sum y)^2}{n}\right)}}$$

Número	Edad(x)	Peso (y)	xy	x ²	y ²
1	7	12	84	49	144
2	6	8	48	36	64
3	8	12	96	64	144
4	5	10	50	25	100
5	6	11	66	36	121
6	9	13	117	81	169
n = 6	∑ x = 41	∑ y = 66	∑ xy = 461	∑ x ² = 291	∑ y ² = 742

Tabla de datos ejemplo 1

Elaborado por Germán Fiallos

Reemplazamos los valores de la tabla en la ecuación de Pearson

$$r = \frac{\sum xy - \frac{\sum x \sum y}{n}}{\sqrt{\left(\sum x^2 - \frac{(\sum x)^2}{n}\right) \left(\sum y^2 - \frac{(\sum y)^2}{n}\right)}}$$

$$r = \frac{461 - \frac{41 \times 66}{6}}{\sqrt{\left[291 - \frac{(41)^2}{6}\right] \left[742 - \frac{(66)^2}{6}\right]}}$$

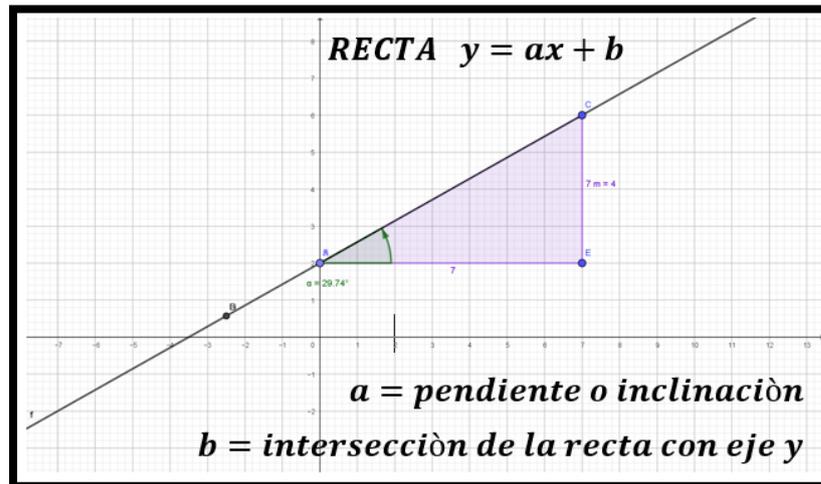
r = + 0.759

Conclusión: analizando la respuesta, podemos inferir:

1. El signo positivo nos indica existe una correlación directa entre las variables peso y edad, esto es, a más edad más peso
2. r = + **0.759** indica que la correlación entre las variables es fuerte

2. 5. Regresiones Lineales: Método de los mínimos cuadrados

El proceso de Regresión nos indica como trazar la línea recta descrita en la correlación y en el gráfico de dispersión, que más se ajuste a los datos entregados; utilizaremos el método de regresión de los mínimos cuadrados para determinar las características de una recta de ajuste del tipo $\hat{y} = ax + b$ de las *variables independiente x, dependiente y*, que pertenecen a una función lineal cuyo gráfico característico es una línea recta,



Elaborado por Germán Fiallos

a. Características del método de los mínimos cuadrados

- Utiliza datos de la variable x para predecir el valor de la variable y
 $y = f(x)$
- a representa la pendiente o inclinación de la recta
- b es el corte o intercepto con el eje y
- Nos indica la razón de cambio de y en función del cambio en los valores de x
- Calcula el modelo matemático $\hat{y} = ax + b$ de una línea recta que proporcione el *mejor trazo o mejor ajuste* para un grupo de datos dispersos
- Utilizando el método de los mínimos cuadrados debemos formular la ecuación de regresión en la forma

$$\hat{y} = ax + b$$

Para lo cual debemos resolver el Sistema de ecuaciones siguiente:

$$\left\{ \begin{array}{l} \sum y = a \sum x + nb \\ \sum xy = a \sum x^2 + b \sum x \end{array} \right.$$

donde a es la pendiente y b es el intercepto con el eje y; son las incógnitas a determinar.

3. RESULTADOS Y DISCUSIÓN

Ejercicio 1.-

Un profesor de estadística realiza un estudio para investigar la relación que existe entre el rendimiento de sus estudiantes en los exámenes y su respectivo nivel de ansiedad. Elige a diez estudiantes de su clase para el experimento. Inmediatamente antes de presentar el examen, los diez estudiantes contestan un cuestionario sobre ansiedad. A continuación, presentamos las calificaciones de la prueba final y los datos de ansiedad para los diez estudiantes

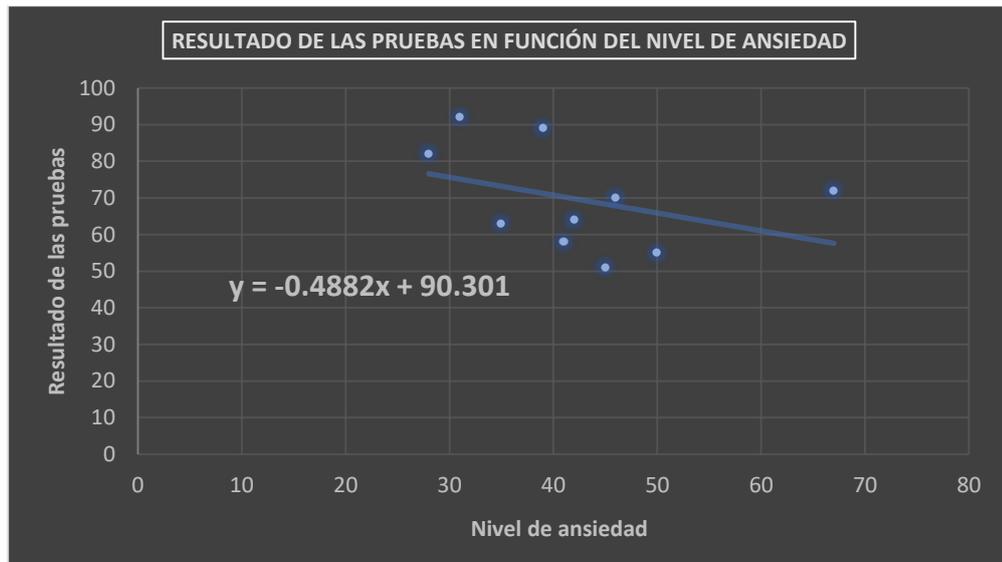
Ansiedad	28	41	35	39	31	42	50	46	45	67
Examen final	82	58	63	89	92	64	55	70	51	72

El estudiante en base a esta información debe cumplir los siguientes objetivos:

- Elaborar un gráfico de dispersión de las parejas de calificaciones. Utilice la ansiedad como variable independiente. Sugerencia: trabaje en Excel
- Suponga que la relación es lineal y determine la intensidad de la correlación entre las variables. Debe indicar si es relación directa o inversa. - Utilice Pearson's en forma analítica y verifique su respuesta con ayuda del Excel
- Determine la ecuación de la recta de mejor ajuste por el método de regresión por mínimos cuadrados, para predecir la calificación del examen final, dado el nivel de ansiedad. - realice el proceso analítico y verifique la respuesta de su ecuación con Excel
- Tomando como base los datos de los diez estudiantes, si un estudiante tiene un nivel de ansiedad de 38, prediga el valor de la calificación que obtendrá en el examen final.

SOLUCIÓN:

- El gráfico de dispersión correspondiente elaborado con ayuda de Excel queda de la siguiente forma:



Elaborado por Germán Fiallos

- b. Para determinar la correlación existente entre las variables completamos la tabla de datos en Excel, datos que serán utilizados para calcular r Pearson's y mínimos cuadrados

	x	y	xy	x^2	y^2
1	28	82	2296	784	6724
2	41	58	2378	1681	3364
3	35	63	2205	1225	3969
4	39	89	3471	1521	7921
5	31	92	2852	961	8464
6	42	64	2688	1764	4096
7	50	55	2750	2500	3025
8	46	70	3220	2116	4900
9	45	51	2295	2025	2601
10	67	72	4824	4489	5184
n = 10	$\sum x = 424$	$\sum y = 696$	$\sum xy = 28979$	$\sum x^2 = 19066$	$\sum y^2 = 50248$

Tabla elaborada por Germán Fiallos

$$r = \frac{\sum xy - \frac{\sum x \cdot \sum y}{n}}{\sqrt{(\sum x^2 - \frac{(\sum x)^2}{n}) \cdot (\sum y^2 - \frac{(\sum y)^2}{n})}}$$

$$r = \frac{28979 - \frac{424 \cdot 696}{10}}{\sqrt{(19066 - \frac{(424)^2}{10})(50248 - \frac{(696)^2}{10})}}$$

$$r = \frac{28979 - 29510,4}{\sqrt{(1088,4)(1806,4)}}$$

$$r = \frac{-531,4}{\sqrt{1\,966\,085,76}}$$

$$r = \frac{-531,4}{1402,17}$$

$$r = -0,3789$$

Análisis de la respuesta:

- El coeficiente r de Pearson es negativo pertenece a una función decreciente
 - Hay relación débil entre las variables
- c. Para regresiones por el método de los mínimos cuadrados tenemos las dos ecuaciones con las cuales debemos encontrar el modelo matemático de la recta de mejor ajuste para las variables x e y , utilizando la tabla de datos anterior

$$\sum x = 424; \quad \sum y = 696; \quad \sum xy = 28979; \quad \sum x^2 = 19066; \quad \sum y^2 = 50248$$

$$\sum y = a \sum x + nb$$

$$\sum xy = a \sum x^2 + b \sum x$$

Reemplazamos los valores:

$$696 = a(424) + 10b$$

$$28979 = a(19066) + b(424)$$

Procedemos a determinar los valores de la pendiente a y el intercepto b con el eje y , para ello resolvemos el sistema de ecuaciones, podemos utilizar cualquier método, sugiero el de sustitución y empezamos eliminando b en las dos ecuaciones

$$696(424) = a(424)(424) + 10(424)b$$

$$28979(-10) = a(19066)(-10) + b(424)(-10)$$

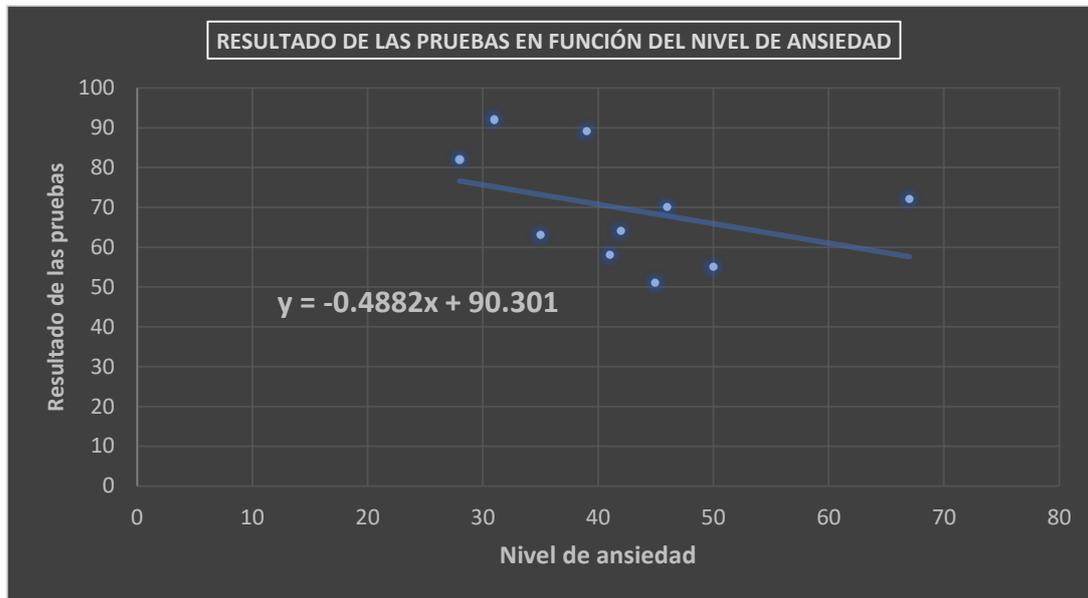
$$295104 = 179776a + 4240b$$

$$-289790 = -190660a - 4240b$$

$$5314 = -10884a$$

$$a = -\frac{5314}{10884}$$

$$a = -0,4882$$



Elaborado por Germán Fiallos

Calculamos el valor de b

$$696 = a(424) + 10b$$

$$696 = (-0,4882)(424) + 10b$$

$$696 = -207.013 + 10b$$

$$696 + 207.013 = 10b$$

$$903,01 = 10b$$

$$b = \frac{903,01}{10}$$

$$b = 90,301$$

Entonces como ya resolvimos el sistema de ecuaciones y encontramos las incógnitas a y b procedemos al reemplazo en el modelo de la ecuación lineal

$$y = ax + b$$

$$y = -0,4882x + 90,301$$

Reemplazamos con las variables originales del problema y obtenemos:

$$\text{Resultado de las pruebas} = -0,4882\text{Nivel de ansiedad} + 90,301$$

Tomando como base los datos de los diez estudiantes, si un estudiante tiene un nivel de ansiedad de 38, prediga el valor de la calificación que obtendrá en el examen final

$$\text{Resultado de las prueba} = -0,4882(38) + 90,301$$

$$\text{Resultado de las prueba} = -18,55 + 90,301$$

$$\text{Resultado de las prueba} = 71,7494/100$$

Ejercicio 2.-

El propietario de una tienda artículos deportivos, lleva un registro de los costos mensuales de publicidad y ganancia (en miles de dólares), los cuales se presentan aquí.

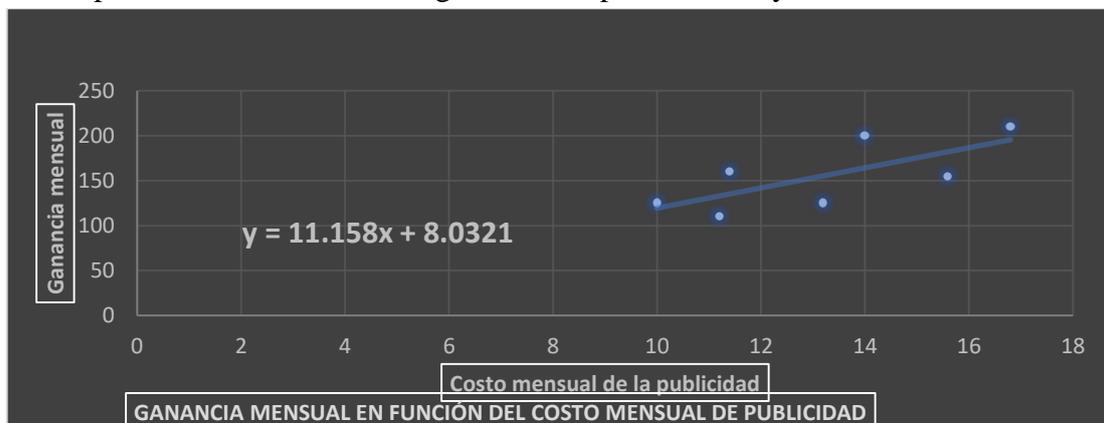
Mes	enero	febrero	Marzo	abril	mayo	junio	julio
<i>Costo mensual de la publicidad</i>	10	14	11,4	15,6	16,8	11,2	13,2
<i>Ganancia mensual</i>	125	200	160	155	210	110	125

El estudiante en base a esta información debe cumplir los siguientes objetivos:

- Realice con ayuda de Excel el diagrama de dispersión correspondiente
- Suponga que la relación es lineal y determine la intensidad de la correlación entre las variables. Debe indicar si es relación directa o inversa. - Utilice Pearson's en forma analítica y verifique su respuesta con ayuda del Excel
- Suponiendo que exista una relación lineal, obtenga analíticamente la recta de regresión por mínimos cuadrados para predecir las ganancias mensuales a partir de los costos mensuales de publicidad. Verifique la respuesta de su ecuación con ayuda de Excel
- En agosto, el gerente planea invertir 17 000 dólares en publicidad. Con base a estos datos, determine la ganancia probable del mes de agosto.

Desarrollo de la solución.

- Empezamos elaborando el diagrama de dispersión con ayuda de Excel



Elaborado por Germán Fiallos

- b. Para el cálculo del coeficiente r de Pearson completamos la tabla de datos cuyos resultados vamos a utilizar. Con el Excel podemos ya obtener el coeficiente r de Pearson sin necesidad de elaborar la tabla mostrada a continuación, pero ahora es necesario para verificar procesos y resultados, luego podríamos optar únicamente por las herramientas de Excel.

	x	y	xy	x^2	y^2
1	10	125	1250	100	15625
2	14	200	2800	196	40000
3	11.4	160	1824	129.96	25600
4	15.6	155	2418	243.36	24025
5	16.8	210	3528	282.24	44100
6	11.2	110	1232	125.44	12100
7	13.2	125	1650	174.24	15625
n = 7	$\sum x = 92,2$	$\sum y = 1085$	$\sum xy = 14702$	$\sum x^2 = 1251,24$	$\sum y^2 = 177075$

- a. Elaboramos el diagrama de dispersión con ayuda del Excel
- b. Suponga que la relación es lineal y determine la intensidad de la correlación entre las variables. Debe indicar si es relación directa o inversa. - Utilice Pearson's en forma analítica y verifique su respuesta con ayuda del Excel. Utilizamos los datos de la tabla anterior

$$r = \frac{\sum xy - \frac{\sum x \cdot \sum y}{n}}{\sqrt{(\sum x^2 - \frac{(\sum x)^2}{n}) \cdot (\sum y^2 - \frac{(\sum y)^2}{n})}}$$

$$\sum x = 92,2; \quad \sum y = 1085; \quad \sum xy = 14702; \quad \sum x^2 = 1251,24; \quad \sum y^2 = 177075$$

$$r = \frac{14702 - \frac{92,2 \cdot 1085}{7}}{\sqrt{\left(1251,24 - \frac{(92,2)^2}{7}\right) \cdot \left(177075 - \frac{(1085)^2}{7}\right)}}$$

$$r = \frac{411}{\sqrt{(36,834) \cdot (8900)}}$$

$$r = \frac{411}{\sqrt{32782,600}}$$

$$r = \frac{411}{572,557}$$

$$r = +0,717$$

Por el signo positivo podemos deducir que entre las variables existe una fuerte correlación directa (FUNCIÓN CRECIENTE), dato que coincide con el obtenido en Excel

c. Suponiendo que exista una relación lineal, obtenga analíticamente la recta de regresión por mínimos cuadrados para predecir las ganancias mensuales a partir de los costos mensuales de publicidad. Verifique la respuesta de su ecuación con ayuda de Excel

a. Ecuación a obtener: $y = ax + b$ con el sistema de ecuaciones

$$\begin{aligned}\sum y &= a \sum x + nb \\ \sum xy &= a \sum x^2 + b \sum x\end{aligned}$$

En estas ecuaciones reemplazamos los datos obtenidos en la tabla de datos del ejercicio

$$\sum x = 92,2; \quad \sum y = 1085; \quad \sum xy = 14702; \quad \sum x^2 = 1251,24; \quad \sum y^2 = 177075$$

El Sistema de ecuaciones queda de la siguiente forma:

$$1085 = a(92,2) + 7b$$

$$14702 = a(1251,24) + b(92,2)$$

Eliminamos b:

$$1085(92,2) = a(92,2)(92,2) + 7(92,2)b$$

$$14702(-7) = a(1251,24)(-7) + b(92,2)(-7)$$

Realizamos las operaciones indicadas

$$100037 = 8500,84a + 645,4b$$

$$-102914 = -8758,68a - 645,4b$$

Sumamos las dos ecuaciones miembro a miembro

$$-2877 = -257,84a$$

$$a = \frac{2877}{257,84}$$

$$a = 11,158$$

Es el valor de la pendiente y lo reemplazamos en una de las dos ecuaciones originales para encontrar b

$$1085 = a(92,2) + 7b$$

$$1085 = (11,158)(92,2) + 7b$$

Realizamos las operaciones y calculamos b

$$1085 = (1028,77) + 7b$$

$$1085 - 1028,77 = +7b$$

$$56,224 = 7b$$

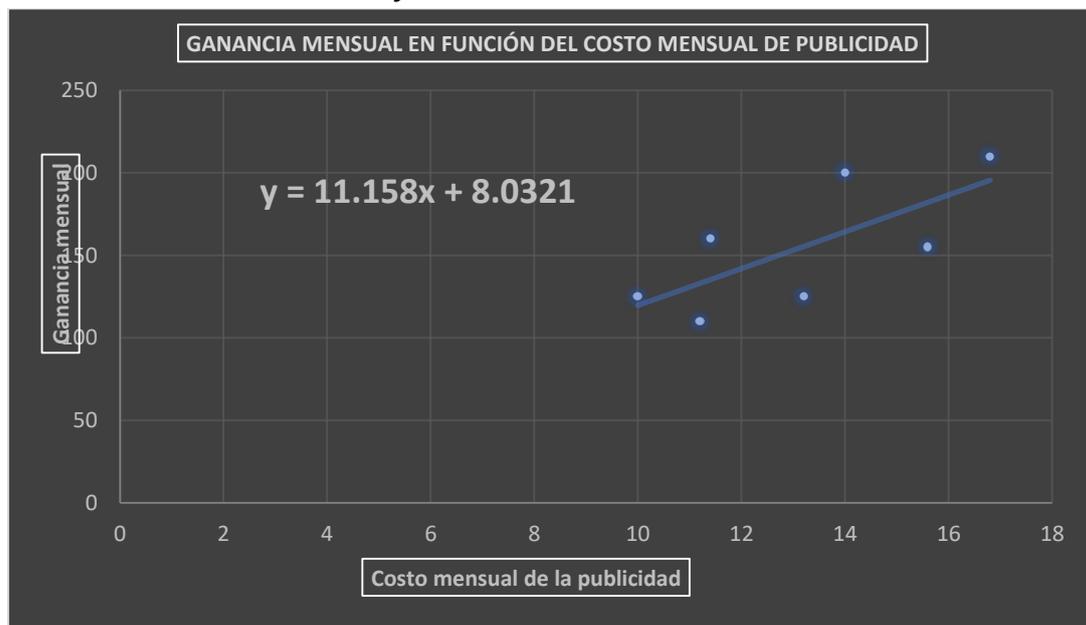
Despejamos b

$$b = \frac{56,224}{7}$$

$$b = 8,0321$$

Por lo tanto nuestra ecuación de regresión es:

$$y = 11,158x + 8,0321$$



Elaborado por Germán Fiallos

Y para nuestros datos se convierte en su equivalente

$$\text{Ganancia mensual} = 11,158 \text{ Costo mensual de publicidad} + 8,0321$$

Volviendo a nuestra pregunta del ejercicio tenemos:

En agosto, el gerente planea invertir 17 000 dólares en publicidad. Con base a estos datos, determine la ganancia probable del mes de agosto

Aplicamos la ecuación de regresión obtenida

$$y = f(x) \Rightarrow \text{Ganancia} = f(\text{publicidad})$$

Ganancia mensual en función del costo mensual de publicidad

$$\text{Ganancia mensual} = 11,158 \text{ Costo mensual de publicidad} + 8,0321$$

$$\text{Ganancia mes de agosto} = 11,158 (17) + 8,0321$$

$$\text{Ganancia mes de agosto} = 189,686 + 8,0321$$

$$\text{Ganancia mes de agosto} = 197,71 \text{ (en miles de dólares)}$$

(Si observa el gráfico realizado en Excel, coincide con la respuesta obtenida por el proceso analítico de regresión por mínimos cuadrados)

4. CONCLUSIÓN:

- Las características analíticas y gráficas de la función lineal se aplican en el estudio de Correlaciones *r de Pearson* y Regresiones lineales por el método de los mínimos cuadrados
- El proceso analítico se verifica con las ecuaciones del Excel

5. LISTA DE REFERENCIAS

- Robert R. Pagano, (2006). Estadística para las ciencias del comportamiento (Séptima edición) Thomson Learning Inc
- Remache, P., Zapata, M., Morales, T., Buenaño, N., Villacís, J., Cáceres, L., Fiallos, G., (2020). Estadística. Publicaciones de la Universidad Indoamérica
- García Ramos, J. A., Ramos Gonzalez, C. D., & Ruiz Garzon, G. (2016). Estadística empresarial. Servicios de Publicaciones de la Universidad de Cádiz.
- Mendenhall, W., Beaver, R., & Beaver, B. (2010). Introducción a la Probabilidad y estadística (Décima ter). Cengage Learning. <http://latinoamerica.cengage.com>
- Johnson, R. A., & Miller, I. (2012). *Miller y Freund probabilidad y estadística para ingenieros* / (8th ed.). Pearson Educación,.
- Levin, R., & Rubin, D. (2004). *Estadística para administración económica*. Pearson Educación, México.
- Oteyza, E. de, Lam, E., Hernández, C., & Carrillo, Á. (2015). *Probabilidad y estadística*. Pearson Educación.
- Ross, S. M., & Valdés Sánchez, T. (2008). *Introducción a la estadística* /. REVERTÉ, S.A.,.
- Torres Huertas, J. (2019). Estadística aplicada a las Ciencias de la salud. Dextra Editorial. <https://elibro.net/es/ereader/utiec/138492?page=78>
- Andrade, R., Gonzalez, E., & Caballero, E. (2014). Un sistema Lógico para el razonamiento y la toma de decisiones: La lógica difusa compensatoria basada en la media geométrica. *Investigación operacional* 32(3), 230-245.
- Cobo, B., & Díaz, C. (2003). MEDIA, MEDIA Y MODA. ¿QUÉ SIGNIFICA ESTO PARA LOS ESTUDIANTES DE SECUNDARIA? *Congreso Nacional de Estadística e Investigación Operativa.*, 27.

- Cortés, F., & Rubalcava, R. (1991). Consideraciones sobre el uso de la estadística en las ciencias sociales. *Estar a la moda o pensar un poco*. Centro de estudios Sociológicos, El Colegio de México, México, DF., 18
- Islas Salomón, C.A. Colin Uribe, M.P. y Morales Téllez, F. (2018). *Probabilidad y Estadística*. Grupo editorial Éxodo.
<https://elibro.net/es/ereader/utieec/128557?page=66>
- Mayen, S., Diaz, C., & Batanero, C. (2009). CONFLICTOS SEMIÓTICOS DE ESTUDIANTES CON EL CONCEPTO DE MEDIANA. *Statistics Education Research Journal*, 8(2), 8.
- Merino, B., & Bernabe, C. (2004). Significado de la media en los libros de texto de secundaria. *Enseñanza de las ciencias: Revista de investigación y experiencias didácticas*, 5, 18.
- Gutiérrez Banegas, A. L. (2020). *Cómo entender estadística fácilmente*. Instituto Mexicano de Contadores públicos.
<https://elibro.net/es/ereader/utiec/130922?page=128>