Universidad Central de Venezuela
Facultad de Humanidades y Educación
Escuela de Educación
Cátedra de Métodos Cuantitativos
Asignatura: Estadística Aplicada a la Educación

Análisis de Correlación y Regresión

Prof: Johnnalid González G.

2018

Objetivo

Entender y evaluar las combinaciones de relaciones entre variables, mediante la comprensión de los conceptos de: asociación, covarianza, correlación, regresión, para analizar sus implicaciones entre ellas, con el fin de aportar mayor información al análisis que se espera en un estudio estadístico.

Ejemplos Previos

En ocasiones, se quisiera analizar si al conjunto de personas se evaluarán sus respectivas edades y pesos, para conocer alguna posible relación.

En educación, los estudiantes que se esfuerzan por realizar todas las actividades, ejercicios, participación en clases, entre otros recursos de aprendizaje, podrían obtener mayores calificaciones, que aquellos que solo se limitan a presentar una evaluación.

A veces, estos planteamientos no tienen el comportamiento esperado, es aquí las técnicas para medir tales realidades.

Coeficiente de relaciones Cuantitativas

1. **Covarianza** → Evaluar la posible relación lineal entre variables cuantitativas.
2. **Coeficiente de Pearson** → Medir la fuerza o intensidad de la relación lineal entre las variables.
3. **Análisis de Regresión** → Estimar la variable dependiente Y basándose en la variación del valor de la variable independiente X.

Covarianza

Es una medida descriptiva que hace referencia a la asociación lineal entre dos variables, también se define como la medida del cambio de una variable que se relaciona uniformemente con un cambio en otra variable de interés.

Un ejemplo del término de covarianza, tenemos el cambio en la actitud hacia las campañas publicitarias de cierta marca juvenil puede variar según el tipo de consumidores para cierto producto nuevo en el mercado. Si se descubren que dos variables cambian juntas sobre una base confiable y uniforme, podemos usar esa información para hacer predicciones así como para las decisiones sobre estrategias de publicidad y marketing.

Covarianza

Esta medida descriptiva se describe de la siguiente manera:

$$\sigma_{xy} = \frac{\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{n-1}$$

Donde:

$(x_i - x)$: desvíos de la variable "X".

$(y_i - y)$: desvíos de la variable "Y".

n – 1: tamaño de la muestra - 1.

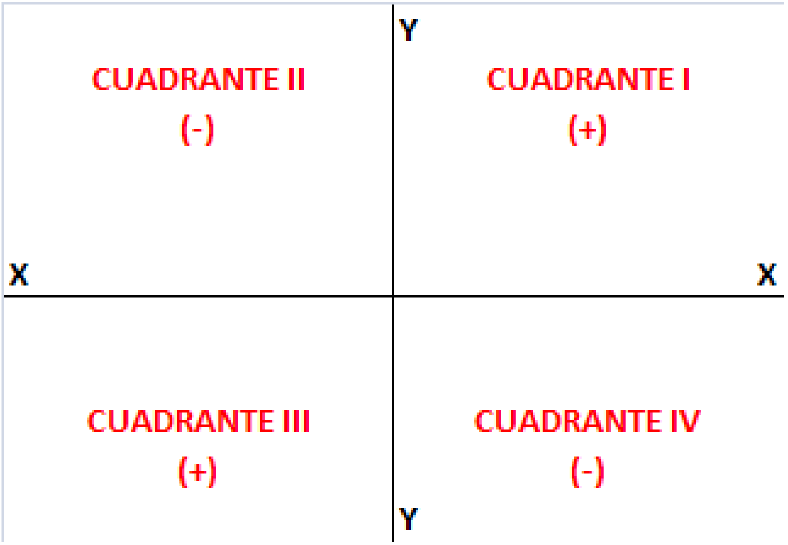
Diagrama de Dispersión

Una forma fácil de describir visualmente la covarianza entre dos variables es haciendo uso de un Diagrama de Dispersión.

Este diagrama grafica la posición relativa de dos variables usando un eje horizontal (eje de las X ó abscisas) y otro vertical (eje de las Y u ordenada) para representar los valores de las respectivas variables. Cuya representación constituyen pares ordenados: (X1, Y1), (X2,Y2), (X3,Y3) y así sucesivamente.

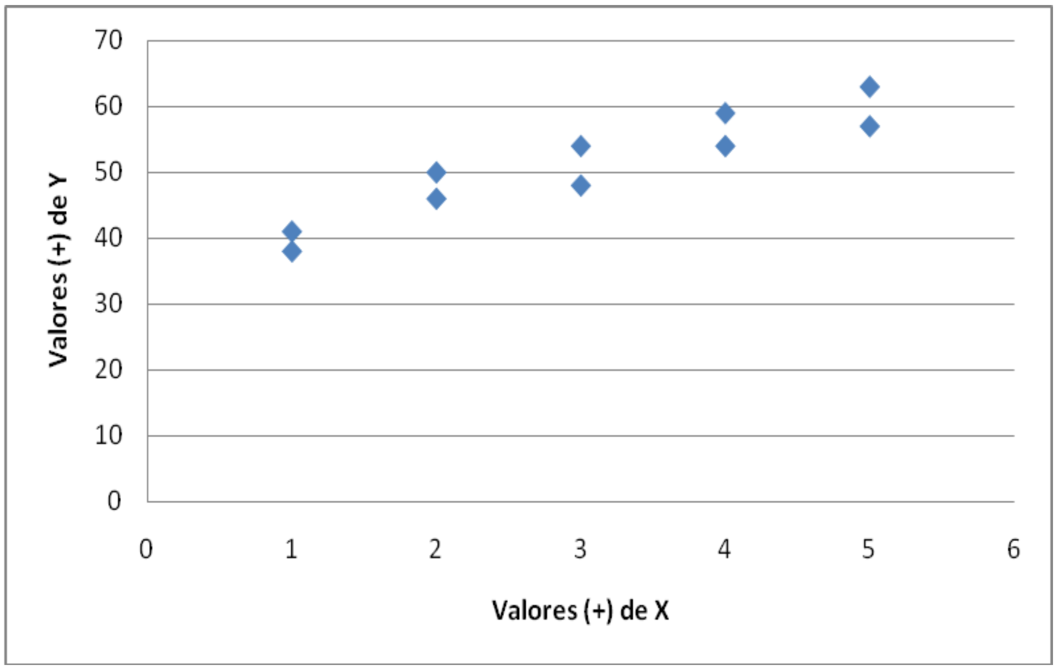
Diagrama de Dispersión

A continuación mostraremos una gráfica donde muestra la ubicación de los signos de los respectivos valores según el grado de asociación existente entre las variables, a los cuales llamaremos *cuadrantes*.



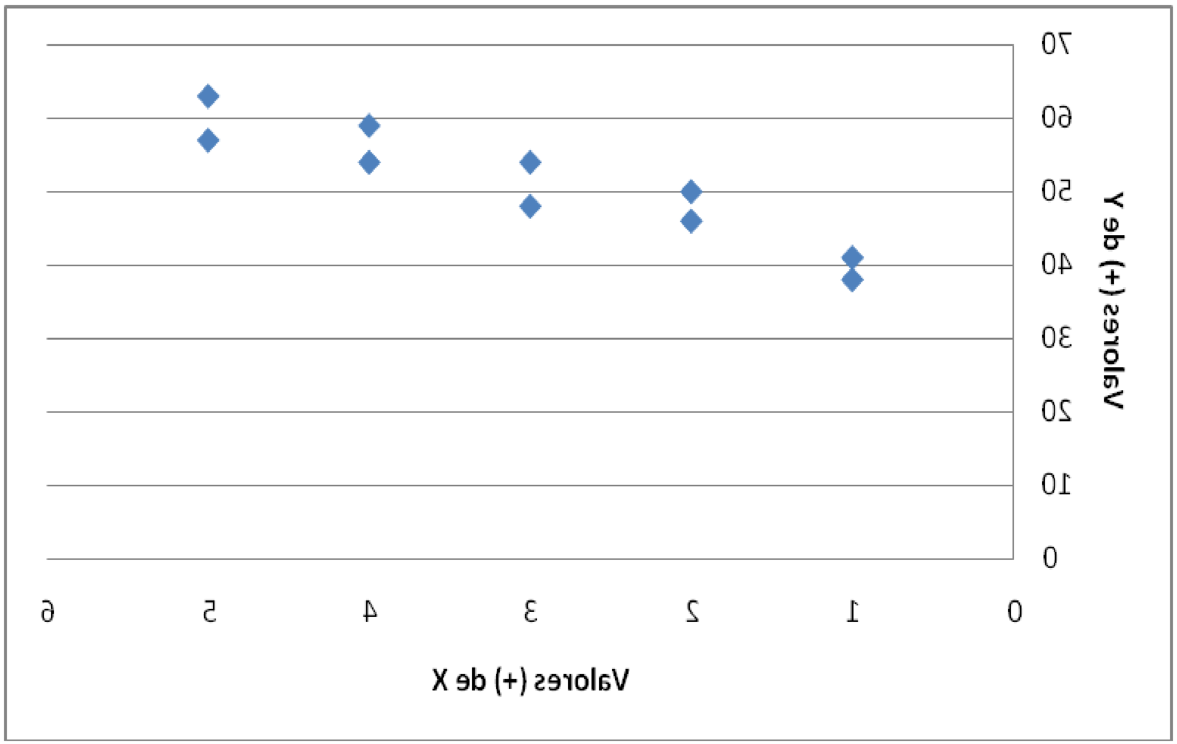
Análisis de la Covarianza

CASO I: Cuando la covarianza es positiva $\sigma_{xy} > 0$



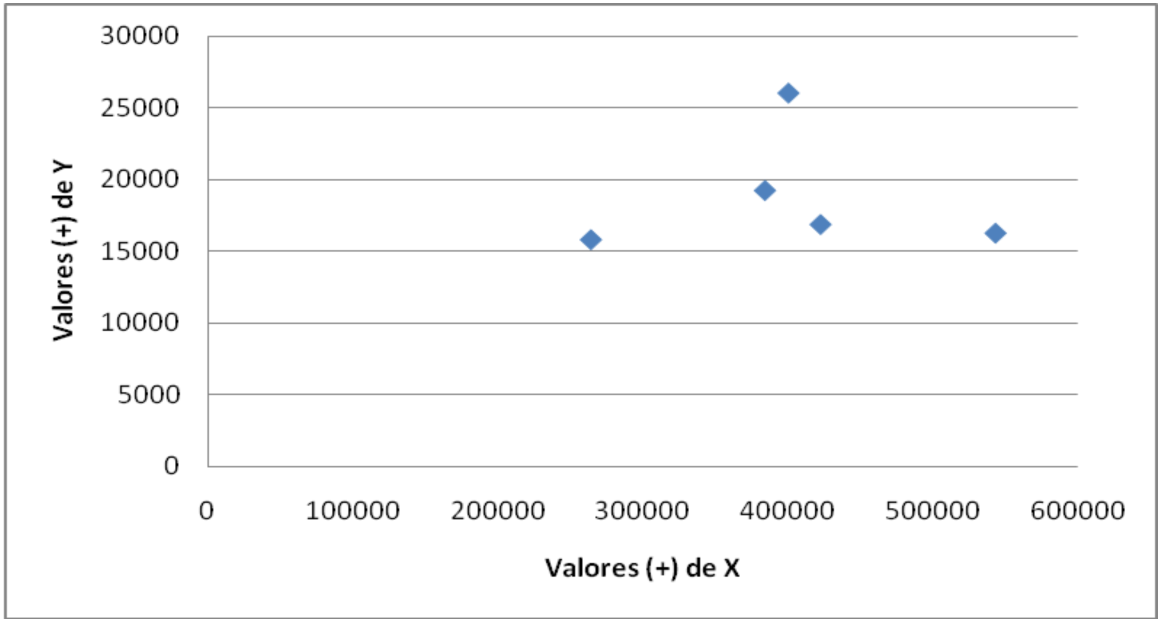
Análisis de la Covarianza

CASO II: Cuando la covarianza es negativa $\sigma_{xy} < 0$



Análisis de la Covarianza

CASO III: Cuando la covarianza es incorrelacionada $\sigma_{xy} = 0$



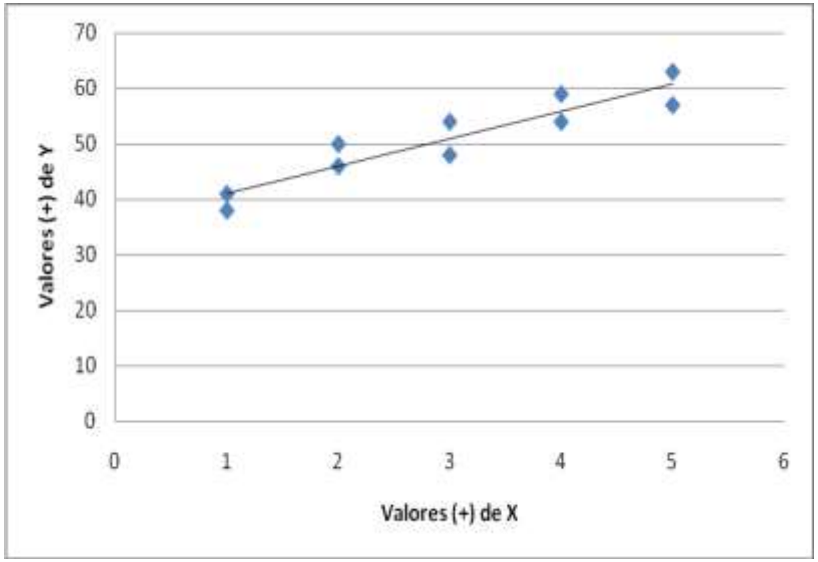
Coeficiente de Correlación de Pearson

Desventaja: Un problema que presenta el uso de la covarianza, resulta que su valor solamente reporta la relación o asociación lineal entre las variables X y Y. Sin embargo, existe una medida estadística que además de indicar que existe relación lineal entre las variables X y Y también permite medir la intensidad o la fuerza de la relación lineal entre las variables, este método se le denomina Coeficiente de Correlación de Pearson.

$$\rho_{xy} = \frac{\sigma_{xy}}{\sigma_x \sigma_y}$$

Interpretación del Coeficiente de Correlación de Pearson

Si todos los puntos de un conjunto de datos caen en una línea recta con pendiente positiva (negativa), el valor del coeficiente de correlación es +1 (-1); esto es, un coeficiente de correlación muestral de +1 (-1) corresponde a una relación lineal positiva (negativa) perfecta entre X y Y.



Tendencia Positiva $p_{xy} > 0$

Conclusión I

- El Coeficiente de Correlación puede valer de –1 < p_{xy}< +1. Los valores cercanos a esos límites indican una fuerte relación lineal. Mientras el coeficiente se acerque más a 0, la relación es más débil.
- Cuando estamos seguros que existe grado de asociación entre dos variables, sea porque conozcamos dichas variables por experiencia o porque así lo indique el coeficiente de correlación previamente calculado, el análisis consistiría en cuantificar la relación existente entre X y Y, con el fin de predecir cuáles serán los valores de una variable, cuando se conocen los valores de la otra variable. En este caso el problema es de un Análisis de Regresión Simple.

Análisis de Regresión

Etimológicamente, la palabra regresión en estadística se utilizó por primera vez en el estudio de variables antropométricas al comparar la estatura de padres e hijos, resultó que los hijos cuyos padres tenían una estatura muy superior al valor medio y tendían a igualarse a éste, mientras, que aquellos cuyos padres eran muy bajos tendían a reducir su diferencia respecto a la estatura media, es decir, "regresaban" al promedio.

En Estadística, el término de regresión radica en la estimación de la variable dependiente Y basándose en la variación del valor de la variable independiente X, es decir, consiste en un procedimiento de estimación o predicción, donde se da lugar a una ecuación matemática que describe dicha relación.

Análisis de Regresión

La finalidad de la ecuación de regresión, es estimar los valores de las variables con base en los valores conocidos de la otra. Tal ecuación lineal se expresa así:

$$Y = a + bX$$

donde:

a: Es el valor estimado de la variable Y cuando la variable X=0.

b: Es el coeficiente de regresión, está expresado en las mismas unidades de la variable Y por cada unidad de la variable X. Su interpretación estaría en función del signo que tome este coeficiente, si “b” es positivo, indica el número de unidades en que varía Y cuando se produce un cambio en una unidad en X (pendiente de la recta de regresión), por el contrario si “b” es negativo, indica la magnitud del decremento en Y por cada unidad de aumento en X.

Ejemplo

Suponga que el gerente de ventas de una red de centro de fotocopiadoras, quien tiene una gran fuerza de venta en cierto negocio de artículos para oficina, quiere determinar si hay alguna relación entre el número de llamadas de ventas hechas en un mes y el número de copiadora vendidas en ese mes.

El gerente selecciona una muestra aleatoria de 10 representantes y determina el número de llamadas de ventas hechas por cada representante el mes pasado y el número de copiadoras que vendió.

Ejemplo

Representante de Venta	Llamadas de Ventas	Copiadoras Vendidas
Miguel	20	30
Rafael	40	60
Alvaro	20	40
Luis	30	60
Susan	10	30
Carlos Ramírez	10	40
Richard	20	40
Elvis	20	50
Andrés	20	30
Juan José	30	70

Ejemplo

Análisis Previos: Parece haber una relación entre el número de llamadas de ventas hechas y el número de copiadoras vendidas. Sin embargo, la relación no es "perfecta" o exacta. Tal es el caso, Juan José hizo menos llamadas que Rafael, pero él vendió más unidades. A continuación definiremos las respectivas variables y los pasos para los cálculos de cada una de las medidas de correlaciones:

a) Definir la variable dependiente e independiente:

Variable Independiente (X): Número de llamadas de ventas.

Variable Dependiente (Y): Número de copiadoras vendidas.

Ejemplo – Análisis de Covarianza

b) Cálculo de la Covarianza: Mide el sentido de la relación entre las variables.

Representantes de Ventas	Llamadas de venta (X)	Copiadoras vendidas (Y)	Xi - X	Yi - Y	(Xi - X) * (Yi - Y)
Miguel	20	30	-2	-15	30
Rafael	40	60	18	15	270
Alvaro	20	40	-2	-5	10
Luis	30	60	8	15	120
Susan	10	30	-12	-15	180
Carlos Ramírez	10	40	-12	-5	60
Richard	20	40	-2	-5	10
Elvis	20	50	-2	5	-10
Andrés	20	30	-2	-15	30
Juan José	30	70	8	25	200
Total	**220**	**450**			**900**

Ejemplo – Análisis de la Covarianza

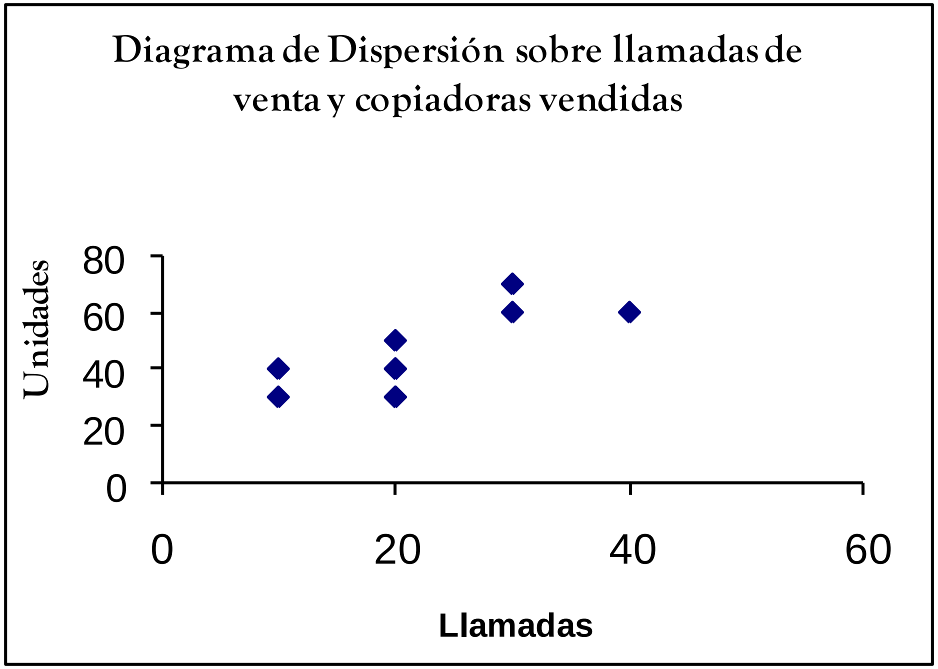
Cálculo de la covarianza:

$$\sigma_{xy} = \frac{\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{n-1} = \frac{900}{9} = +100$$

Como $\sigma_{xy} > 0$, entonces decimos que existe una correlación positiva entre las variables, es decir, que a medida que aumenta las llamadas realizadas por los representantes de ventas aumenta las ventas de copiadoras. Tal fenómeno se puede apreciar mediante el siguiente Diagrama de Dispersión de las variables respectivas.

Ejemplo – Análisis de la Covarianza

Tal situación, se puede apreciar mediante el siguiente Diagrama de Dispersión de las variables respectivas.



Ejemplo – Análisis de la Correlación

Por lo tanto, el coeficiente de correlación de Pearson sería:

$$\rho_{xy} = \frac{\sigma_{xy}}{\sigma_x * \sigma_y} = \frac{100}{(8,72)*(13,60)} = +0,76$$

Como ρ_{xy} = 0.759, por ser un resultado positivo, decimos que hay una relación directa entre el número de llamadas de venta y el número de copiadoras vendidas, relación observada en el diagrama de dispersión. El resultado de dicho coeficiente es de 0.759, indica una fuerte relación entre las variables (esta cercano a +1), es decir, un aumento de un 25% en las llamadas llevaría probablemente a un aumento de un 25% en las ventas.

Ejemplo – Análisis de Regresión

Consiste en estimar los valores de la ecuación de regresión lineal, tal y como sigue:

Y = a + bX

Aplicamos las fórmulas de "a" y "b" que son los parámetros a estimar en la recta de regresión.

$$b=\frac{\sum XY-n\bar{x}\,\bar{y}}{\sum X^2-n\bar{x}^2}=\frac{10.800-10*(22)(45)}{(5.600)-10*(22)^2}=1{,}1842$$

y,

$$a=\frac{\sum Y}{n}-b\frac{\sum X}{n}=\bar{Y}-b\,\bar{X}=\frac{45}{10}-(1.1842)\frac{22}{10}=18{,}947$$

Ejemplo – Análisis de Regresión

La ecuación de regresión estimada: $\hat{Y} = a + bx_i = 18.9476 + 1.1842X$

Interpretaciones de "a" y "b":

El valor de b = 1.1842, significa que por cada llamada más, el vendedor puede esperar un aumento de 1.2 en el número de copiadoras vendidas.

El valor de a = 18.9476, significa que si no se hace ninguna llamada (X = 0), se venderán 18.9476 copiadoras, y por lo tanto, no se deberá usar para estimar el número de copiadoras que se venderán.

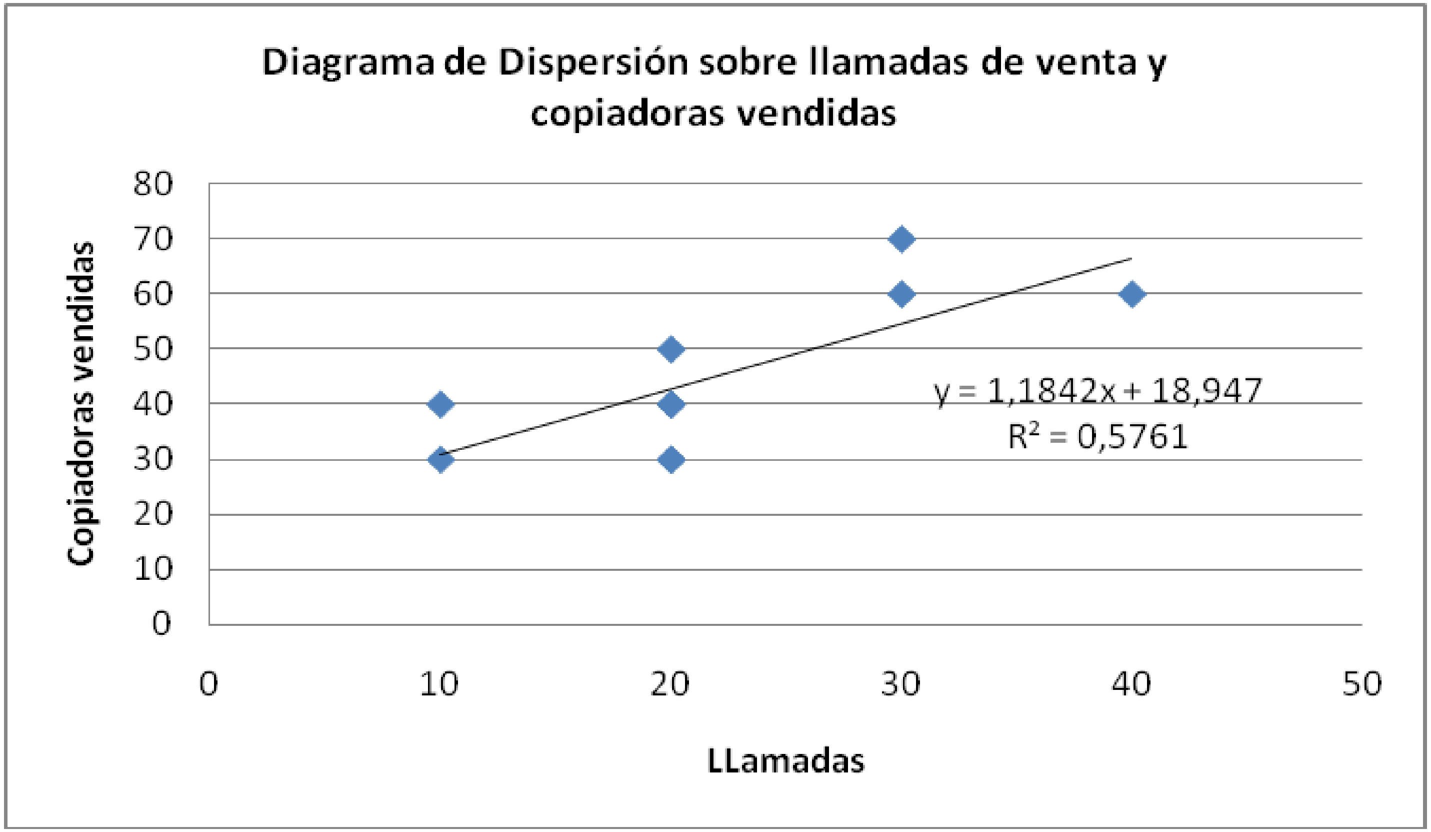
Ejemplo – Análisis de Regresión

Para ello, se mostrarán los registros que se esperaría en las ventas estimadas por cada representante de venta, tal y como sigue:

Representantes de Ventas	**Llamadas de Ventas (X)**	**Ventas Estimadas (Y´)**
Miguel	20	42,6316
Rafael	40	66,3156
Alvaro	20	42,6316
Luis	30	54,4736
Susan	10	30,7896
Carlos	10	30,7896
Richard	20	42,6316
Elvis	20	42,6316
Andrés	20	42,6316
Juan	30	54,4736
Total	**220**	

Ejemplo – Análisis de Regresión

También se puede observar en el diagrama de dispersión de la recta de regresión calculada:



Ejemplo – Coeficiente de Determinación

Siguiendo con el ejemplo, para conocer la calidad de la línea de tendencia calculada analicemos el coeficiente de determinación, cuyo análisis es el siguiente:

d) Coeficiente de Determinación: Mide la calidad del ajuste de la recta de regresión.

r^2 = 0.576, esto indica que el 57.6% de la variación en el número de copiadoras vendidas, se explica, o se debe a, la variación en el número de llamadas de ventas.

En resumen, el análisis de regresión y correlación son dos técnicas estrechamente relacionada y comprenden una forma de estimación, es decir ambas técnicas comprende el análisis de los datos muestrales para saber qué es y cómo se relacionan entre sí dos o más variables en una población.