La **Regresion Lineal**, es una tecnica estadistica, para predecir o estimar, una variable cuantitativa en funcion de otra variable cuantitativa.

La primera seria **Y**, la segunda seria **X**, la **Y** es la dependiente, ya que depende del valor que le demos a **X**, y la **X** es nuestra variable independiente, esta explica a **Y.**

**Y,** es aquella que necesitamos predecir o estimar

**X,** es nuestra variable explicativa

Consiste en modelar una ecuacion de la recta, recordando calculo diferencial, esta puede ser descendente o ascendete.

**Ejemplo:**

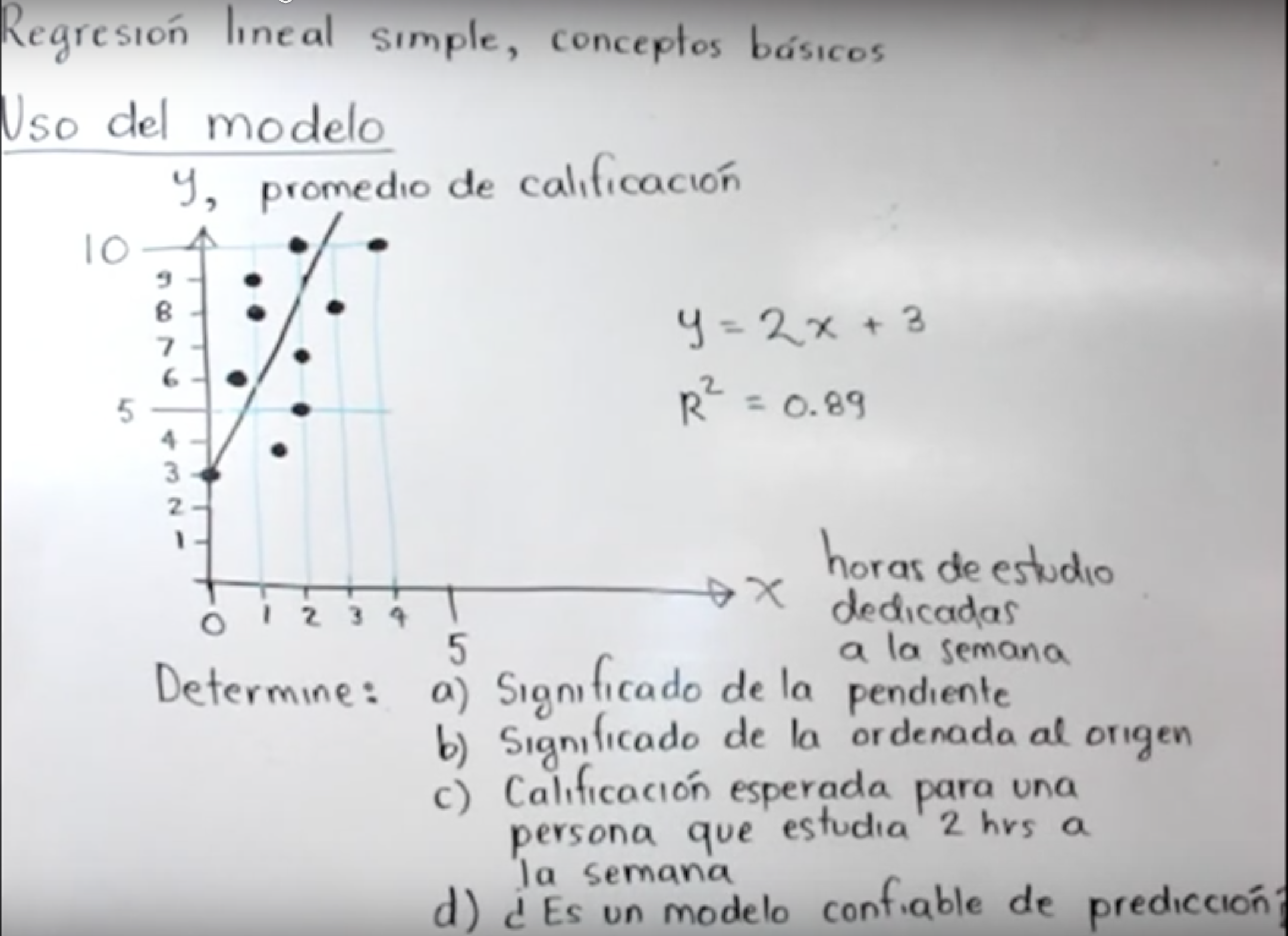
**Predecir el sueldo mensual en funcion a la edad de un trabajador**

**Y <- Sueldo** DEPENDIENTE

**X <- Edad** INDEPENDIENTE

Sii la edad sube (x), el sueldo sube (y), y viceversa…

**Conclusión:** Si ambas variables aumentan o disminuyen entonces tienen una relación xdirecta.



1. **Significado de la pendiente**

Podemos decir que, si **X** son las horas de estudio y **Y,**  son el promedio de calificación obtenido, entonces por cada hora adicional o incremento unitario, la calificación aumenta dos puntos y = 2x + 3

1. **Significado de la ordenada al origen**

Que una persona que no estudie (0) horas, se espera que saque en promedio 3 puntos. ya que si 2 (0) + 3 = **3**

1. **Calificación esperada para una persona que estudia 2 horas a la semana**

Simplemente sustituimos, y ponemos Y = 2 (2) + 3 = **7**

1. **Es un modelo confiable de predicción?**

Se dice que el modelo es confiable ya que el coeficiente de determinación R^2, se acerca mas a 1, ya que es 0.89

**Ejemplo:**

Deseamos predecir el tiempo que demora una constructora en construir un condominio en funcion al numeo de trabajadores que contrato de realizar la obra

**Y,** tiempo de construccion, a predecir

**X**, numero de trabajadores, explicativa

**Coclusion:** Mientras mas trabajadores contratan, menos tiempo se demorara el tiempo de construccion. Y mientras menos trabajadores, mas tiempo se demorara el tiempo.

X y Y tienen una relacion inversa, que es lo contrario a una relacion directa…

**Ejemplo:**

Predecir la nota de un alumno, en función a la cantidad de horas a la semana que estudio. Tenemos de muestra seis (6) alumnos:

|  |  |
| --- | --- |
| **Hrs de estudio (X)** | **Nota (Y)** |
| 3 | 8 |
| 6 | 10 |
| 8 | 15 |
| 2 | 8 |
| 1 | 5 |
| 6 | 12 |

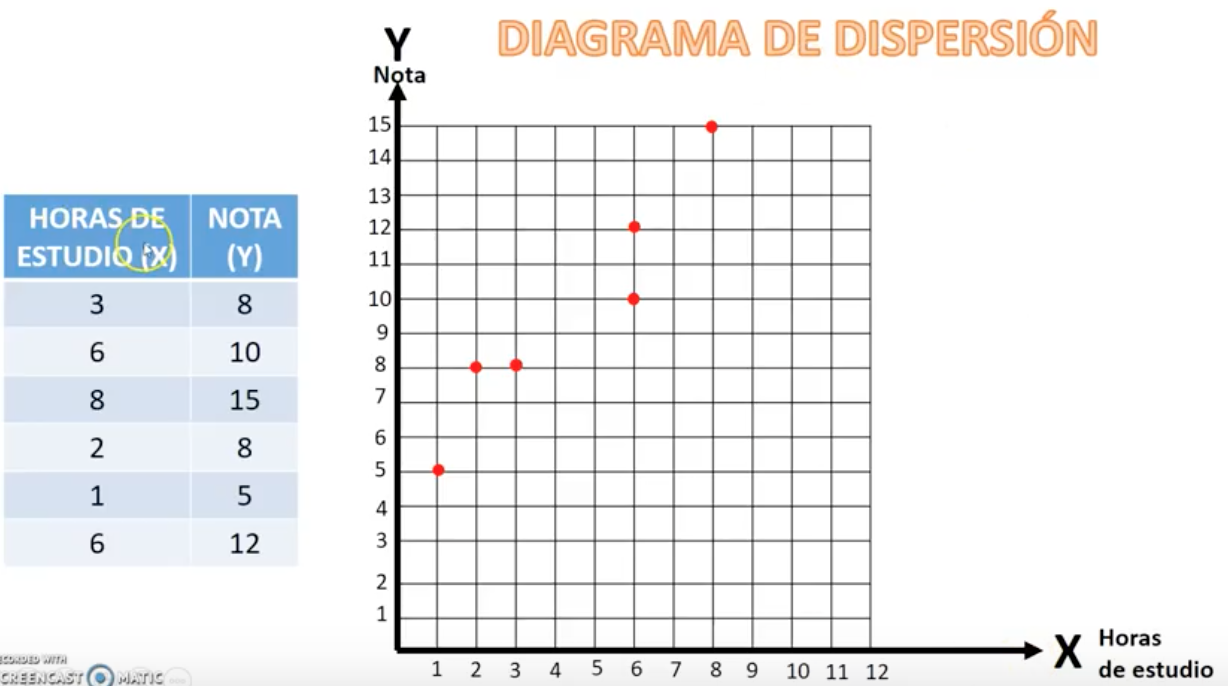
**IMPORTANTE:** Lo ideal a realizar, es un **diagrama de dispersion,** el mismo nos ayudara a ver la relacion entre las variables **X** y **Y.**

**El Diagrama de dispersión,** no es mas que graficar ambas variables en un plano cartesiano, mediante puntos de coordenadas.

X <- horas de estudio, explicativa

Y <- nota, a predecir

Se establece una **relación directa**, ya que a medida que la X (horas de estudio) aumenta, entonces la Nota (Y), va aumentando de igual manera...

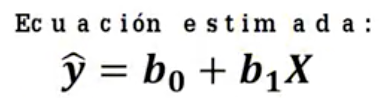


La tecnica de regresion lineal, consiste en modelar una linea recta que una los puntos; no existe una recta como tal que pase por cada punto, deberian estar alineados los puntos…

Podemos decir que se puede intentar graficar, una linea donde se abarque la mayor cantidad de puntos. **Existen infinitas rectas posibles que pasan cerca de los 6 puntos.**

Pero como todo en la vida, existe una manera mas optima de hacerla, esta es la **Ecuacion Estimada o ÿ (**y sombrero**)...**

La Ecuacion para sacar la linea recta, mediante la ecuación estimada:



La mejor recta, se define como aquella que logra disminuir las distancias entre los puntos y la recta como tal que se modela…

La Ecuación estimada Y, se obtiene calculado los coeficientes b0 y b1 con el **metodo de minimos cuadrados…**

Pero quienes son los coeficientes b0 y b1? quienes son estas constantes ???

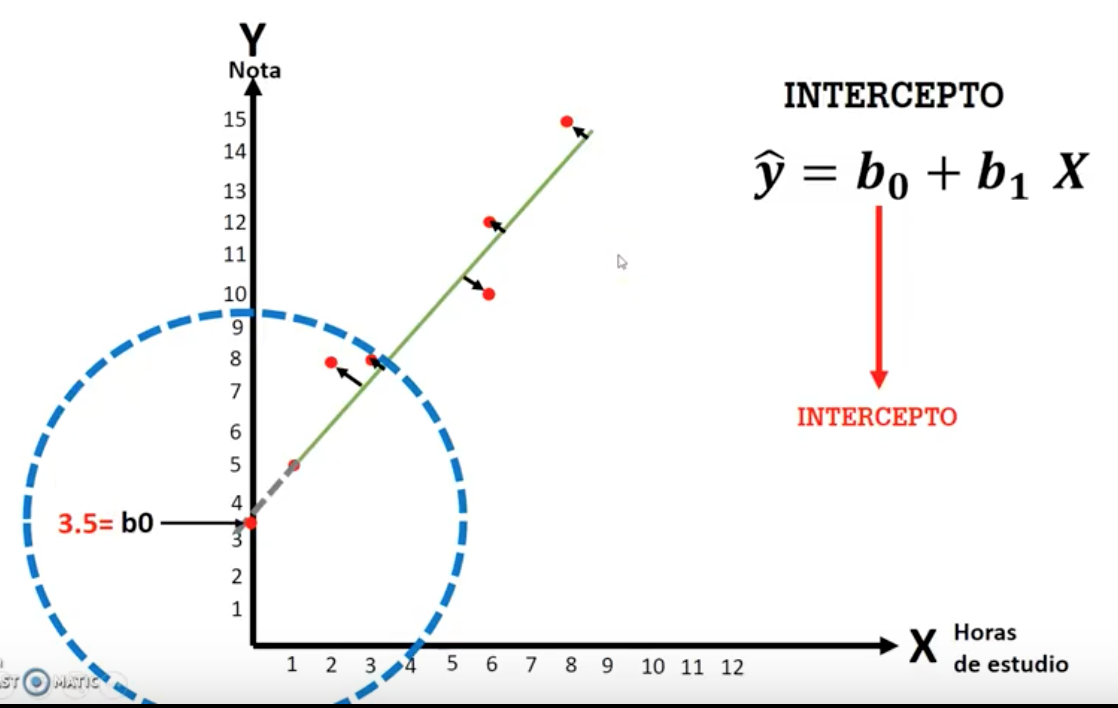
b0 se llama intercepto

b1 se llama pendiente

En la ecuacion de la recta lo conocemos mejor como:

**Y = b + mX**

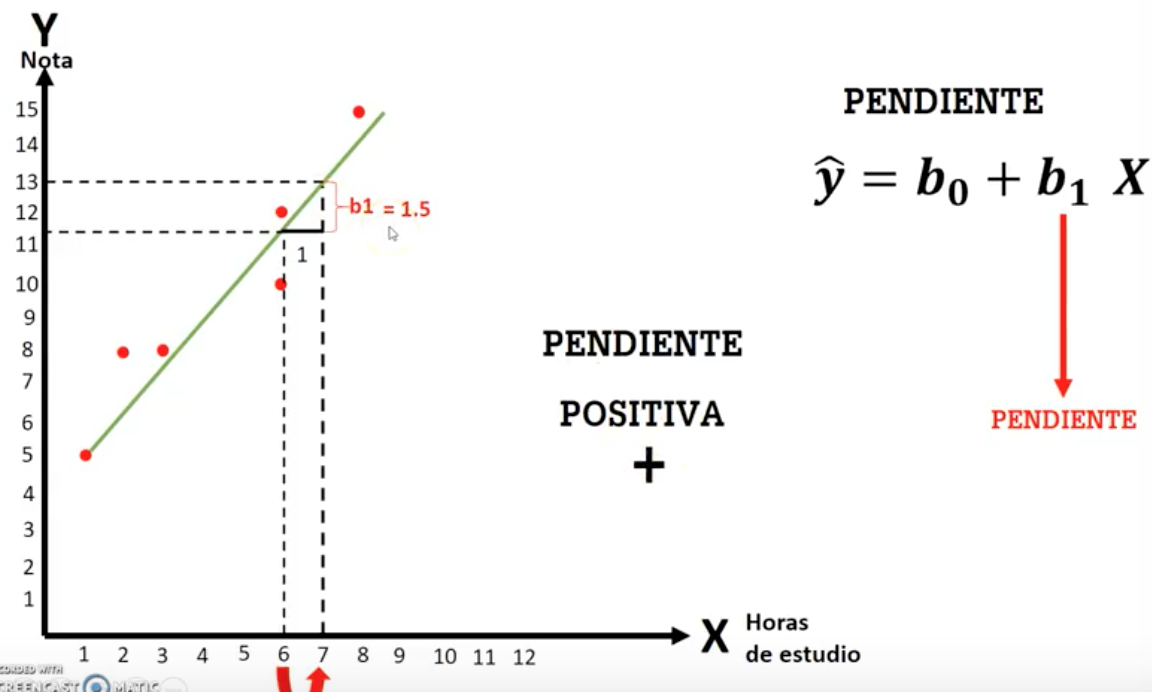
Donde b es el intercepto y m es la pendiente de la recta,

b0 o b, viene siendo el punto el cual donde la recta intercepta al eje **Y**

Si la recta no da como esa, se alarga un poco mas,

Ahora bien la pendiente (m) o b1, indica la inclinacion de nuestra recta con respecto al eje horizontal (X). A mayor pendiente, mayor es su valor, y mas positiva, si la pendiente es hacia abajo, sera negativa, horizontal es 0 y vertical totalmente se conoce como asintota y es indeterminada.

En nuestra gráfica, la pendiente es positiva, escogemos en una distancia de 1 unidad, por ejemplo de **6 a 7,**  cualquier numeros elegidos funciona, el triangulo formado, la altura del mismo, es decir el cateto opuesto, es la pendiente, se mide hasta donde llega en el eje Y, y se calcula la distancia entre los puntos, y esa es nuestra pendiente m (b1):



La ecuacion estimada es:

**Y = 3.5 + 1.5X**

Con este metodo, buscamos es estimar la variable Y utilizando esta ecuacion a partir de la **X.**

Si estudia 0 hora = da 3.5 de nota

Si estudia 1 hora = da 5 de nota

Si estudia 2 horas = da 6.5 de nota

Si estudia 4 horas = 9.5

Se nota un patron y es que la nota aumenta en 1.5, que es nuestra pendiente

**Conclusión: por cada hora adicional de estudio, la nota del alumno aumenta en 1.5 puntos**