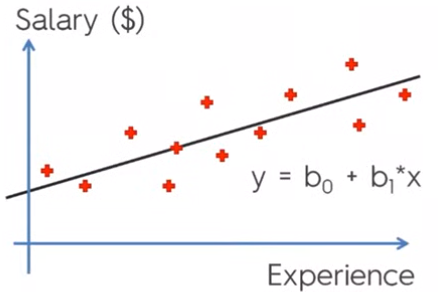
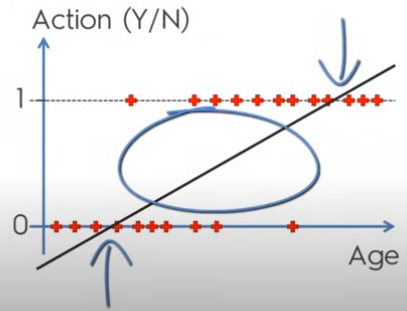
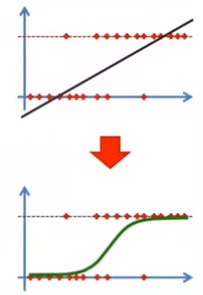
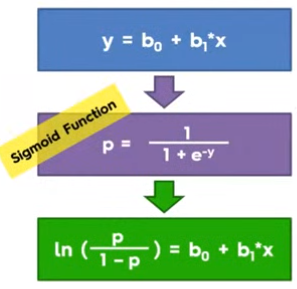
**¿Por qué elegimos Regresión Logística?**

****

**Tenemos el salario vs la experiencia de un trabajador, si vemos el grafico, la recta no cubre nuestro espectro de datos y muchas veces la respuesta de estos datos es 0 o1 representando un si o un no.**

****

**Entonces una regresión de este tipo no podría cubrir el comportamiento de los datos de ninguna forma, es la característica del DATASET que vamos a trabajar.**

**Entonces en la ecuación de la línea recta podríamos considerar un senoide a travez de la función sigmoid dada por una ecuación de probabilidad, probabilidad de que se obtenga un número y si nuestros valores van entre 0 y 1, entonces se combinan las ecuaciones y se obtiene una curva senoidal que define la probabilidad de que los valores pertenezcan a un grupo o al otro.**

**REGRESIÓN LOGÍSTICA SIMPLE**

Este modelo tiene la forma:



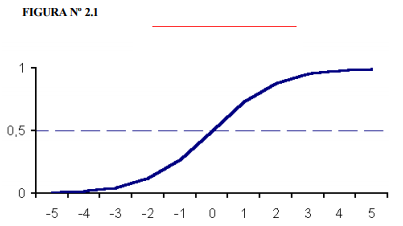
De esto se deduce que:



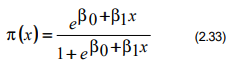
Por tanto i ε , no puede tener distribución normal debido a que toma valores discretos, el Modelo de Regresión Lineal Simple, no es aplicable para el caso de variable respuesta de tipo dicotómico. En el Análisis de Regresión Lineal simple, el punto inicial del proceso de estimación del modelo es un gráfico de dispersión de la variable respuesta versus la regresora, pero este gráfico resulta limitado cuando sólo hay dos valores posibles para la variable respuesta, por tanto se debe usar otros gráficos, éstos resultan de la suavización de los valores de la variable respuesta, representando después los valores de la variable respuesta versus la regresora. La notación que se usará en el presente trabajo para la Regresión Logística es misma que emplea Hosmer y Lemeshow (2000).



Que representa la media condicional de y = 1dado x , donde π ( ) x representa la probabilidad de que ocurra y = 1, ciertamente no se espera que (2.32) tenga relación lineal dentro del rango de la variable regresora. ¿Qué hay de la relación entre π (x ) y x para valores intermedios de x ? Se espera una relación curvilínea. Para cualquier valor grande de x , π ( ) x tomará valores cercanos a 1 y para valores pequeños de x , π ( ) x tomará valores cercanos a cero. El gráfico que muestra el comportamiento de π (x ) versus x es:



curva en forma de S o sigmoide que tiene las propiedades requeridas para π (x ) y que tiene las propiedades de una función de distribución de probabilidad acumulada, para esta probabilidad se usa la función de distribución acumulada de la distribución logística dada por:

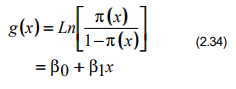


(2.33) tiene un gráfico similar a la Figura Nº 2.1, cuando β0 < 0 y β1 > 0, además este modelo toma valores en el intervalo [0,1 ]. Cuando ,P[y] = 1 = 5.0 el valor de x es:



que brinda información muy útil.

Una transformación de π ( x ) que es lo central del estudio de la Regresión Logística es la transformación logit, transformación que se define en términos π ( x ) como:



Lo importante de esta transformación es que tiene muchas propiedades semejantes al Modelo de Regresión Lineal simple, por ejemplo es lineal en sus parámetros, puede ser continua y puede tomar cualquier valor real dependiendo de x . Para el Modelo de Regresión Lineal simple, la variable respuesta, de (2.4) se expresa como:



para la variable respuesta dicotómica lo expresamos como:



veamos que ocurre con este modelo:



Entonces εi tiene distribución binomial con media cero y varianza:



Por tanto la distribución condicional de la variable respuesta tiene distribución de probabilidad binomial con media π ( x ) .

El lado izquierdo de (2.34) se llama también logaritmo de ODDS RATIO o razón de probabilidades de y = 1contra y = 0 , específicamente:



o también llamado razón de ventaja a favor de éxito.

# Bibliografía

Ca. (s.f.).

Salcedo Poma, C. M. (s.f.). *sisbib.unmsm*. Obtenido de https://bit.ly/3oWD7jX