

Unos pocos conceptos básicos de Regresión Lineal que necesitáis saber

DATOS = MODELO + ERROR

$$Y_i = \hat{Y}_i + (Y_i - \hat{Y}_i)$$

Donde  $\hat{Y}$  se refiere a la ecuación de regresión. En el caso de regresión lineal simple:

$$\hat{Y} = A + B X$$

A (ordenada en origen) y B (pendiente) se obtienen minimizando la suma de  $(Y_i - \hat{Y}_i)^2$  (criterio de mínimos cuadrados)

$$Y_i = \hat{Y}_i + (Y_i - \hat{Y}_i)$$

$$s^2_y = s^2_{\hat{y}} + s^2_e$$

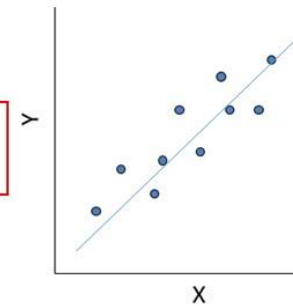
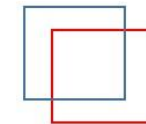
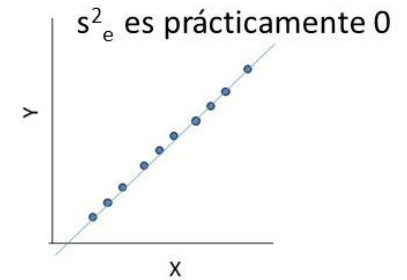
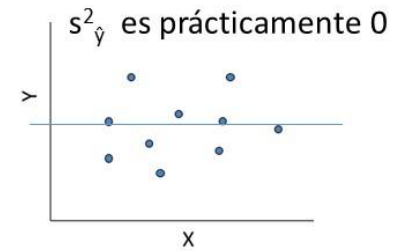
$s^2_y$  es la varianza (total) de Y

$s^2_{\hat{y}}$  es la varianza de Y que puede ser explicada por la ecuación de regresión

$s^2_e$  es la varianza de Y que *no puede* ser explicada por la ecuación de regresión

**¿Cuán bueno es el modelo?**

$r^2$  es la proporción de varianza de Y que puede explicar la ecuación de regresión. Este coeficiente se le suele llamar "coeficiente de determinación"



-- $R^2$  indica cuánta varianza de la variable dependiente puede ser explicada por (o está asociada con) la ecuación de regresión compuesta por uno o más predictores

--en regresión lineal simple,  $r^2$  es el cuadrado del coeficiente de Pearson; observa también que la pendiente de la línea de regresión en puntuaciones típicas siempre coincide con el coeficiente de Pearson.

--La ecuación de regresión proporciona los coeficientes en las puntuaciones no estandarizadas ( $b$ ) y en las puntuaciones estandarizadas ( $\beta$ ). Para saber cuáles son los mejores/peores predictores, hay que comprobar los coeficientes estandarizados (estos coeficientes de regresión están entre -1 y +1, y dan una idea de la relación de cada predictor con la variable dependiente; ten en cuenta que las puntuaciones estandarizadas no tienen unidades, por lo que puedes comparar diferentes predictores).

--Si hay varios predictores (variables independientes, en SPSS), entonces la situación ideal es que estos predictores no estén relacionados entre sí. Si hay una alta relación entre los predictores, podemos sufrir problemas de colinealidad - puede que quieras comprobar la colinealidad en SPSS o JASP (por ejemplo, mirando el Factor de Inflación de la Varianza [VIF en inglés: Variance Inflation Factor]; ¿está por encima de 10 o no? Este es un criterio muy de máximos, porque idealmente debería ser 1, no mucho mayor)

--De forma predeterminada, introducimos todos los predictores (variables independientes, en SPSS) en la ecuación de regresión. Pero hay otras opciones. Por ejemplo, en el procedimiento "por pasos" ("stepwise"), sólo los predictores mejores (de acuerdo con un criterio) entran en la ecuación de regresión (es decir, ¿qué sentido tiene incluir predictores que básicamente no ayudan a predecir la variable dependiente? Es la navaja de Occam o principio de parsimonia)