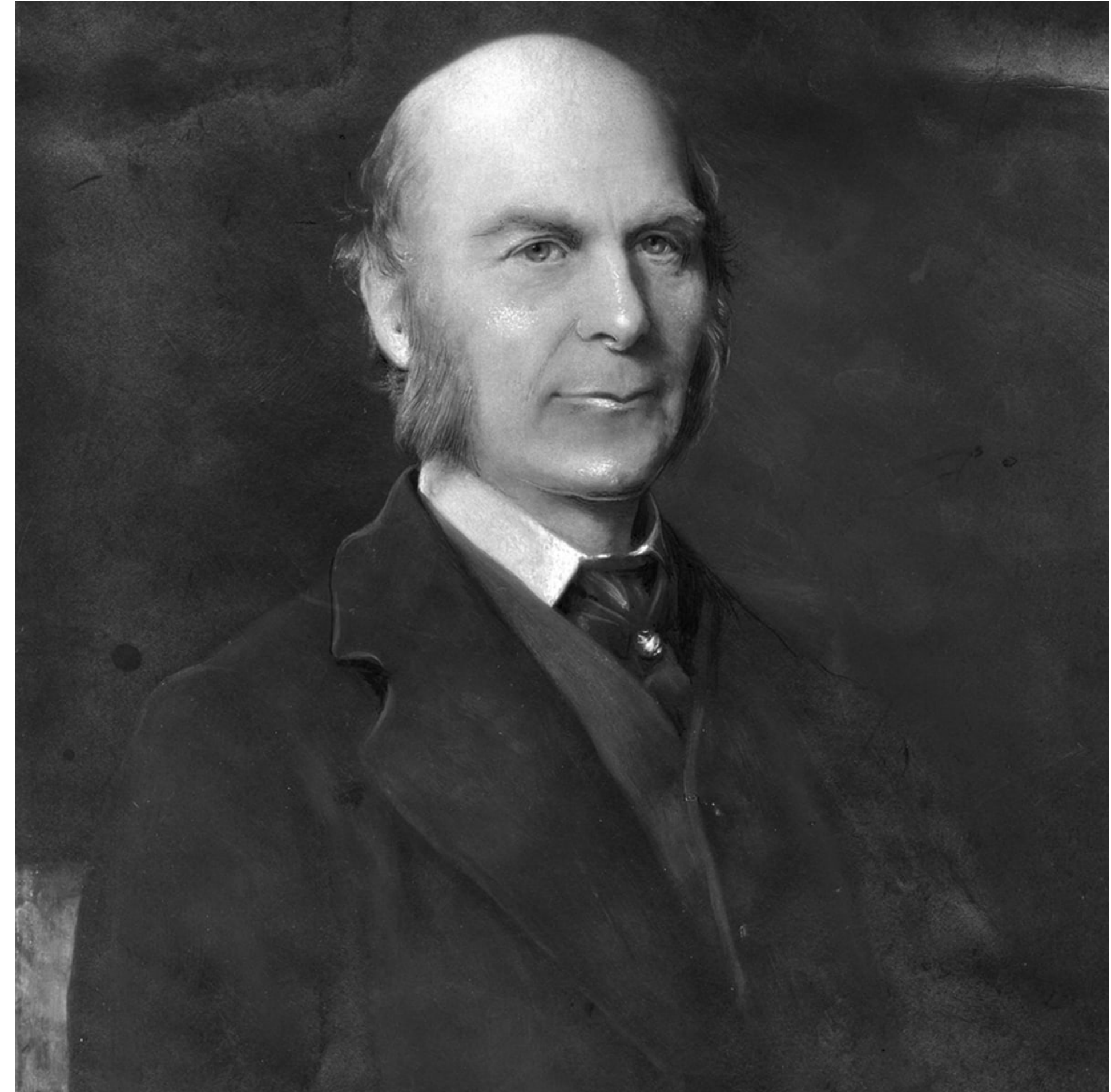


Intuición sobre Regresión Lineal

Luis R. Furlán (Julio 2021)

Historia

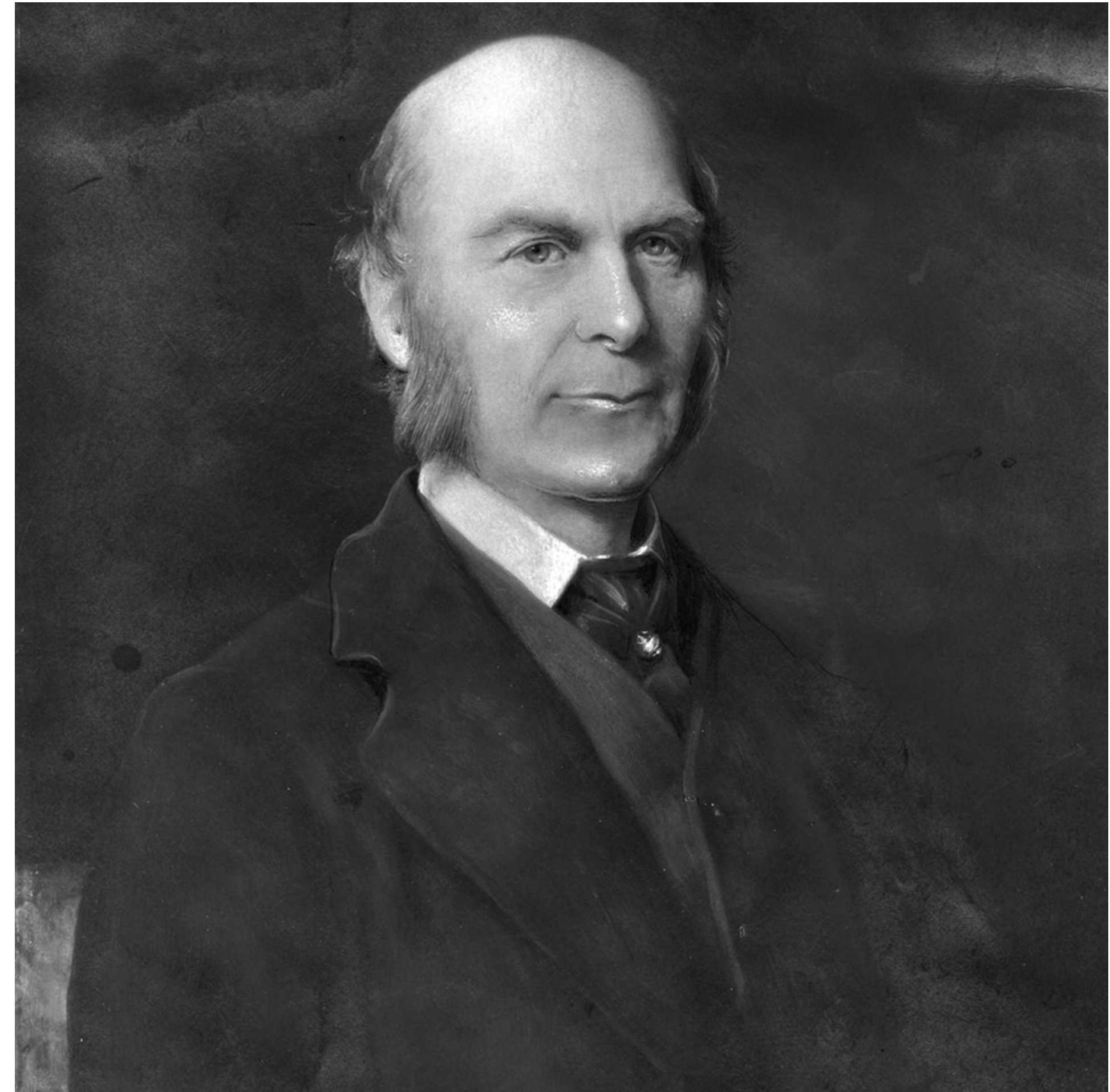
En los años 1800s, Francis Galton estudiaba la relación entre los padres y los hijos. En particular, investigó la relación entre las estaturas de los padres y sus hijos.



Historia

Lo que descubrió fue que el hijo de un hombre tiende a ser aproximadamente tan alto como su padre.

Sin embargo, su mayor aporte fue encontrar que la estatura del hijo tiene la tendencia de ser más cercana al promedio general de la estatura de toda la gente.



Ejemplo

Shaquille O'Neal es realmente alto: 2.2 metros.

Si Shaq tiene un hijo, la probabilidad es que su hijo será bastante alto también. Sin embargo, la estatura de Shaq es tan atípica que también hay buena probabilidad de que su hijo no sea tan alto como él.



Ejemplo

Este es el caso. El hijo de Shaq es bastante alto (2.0066 m), pero no tanto como su padre.

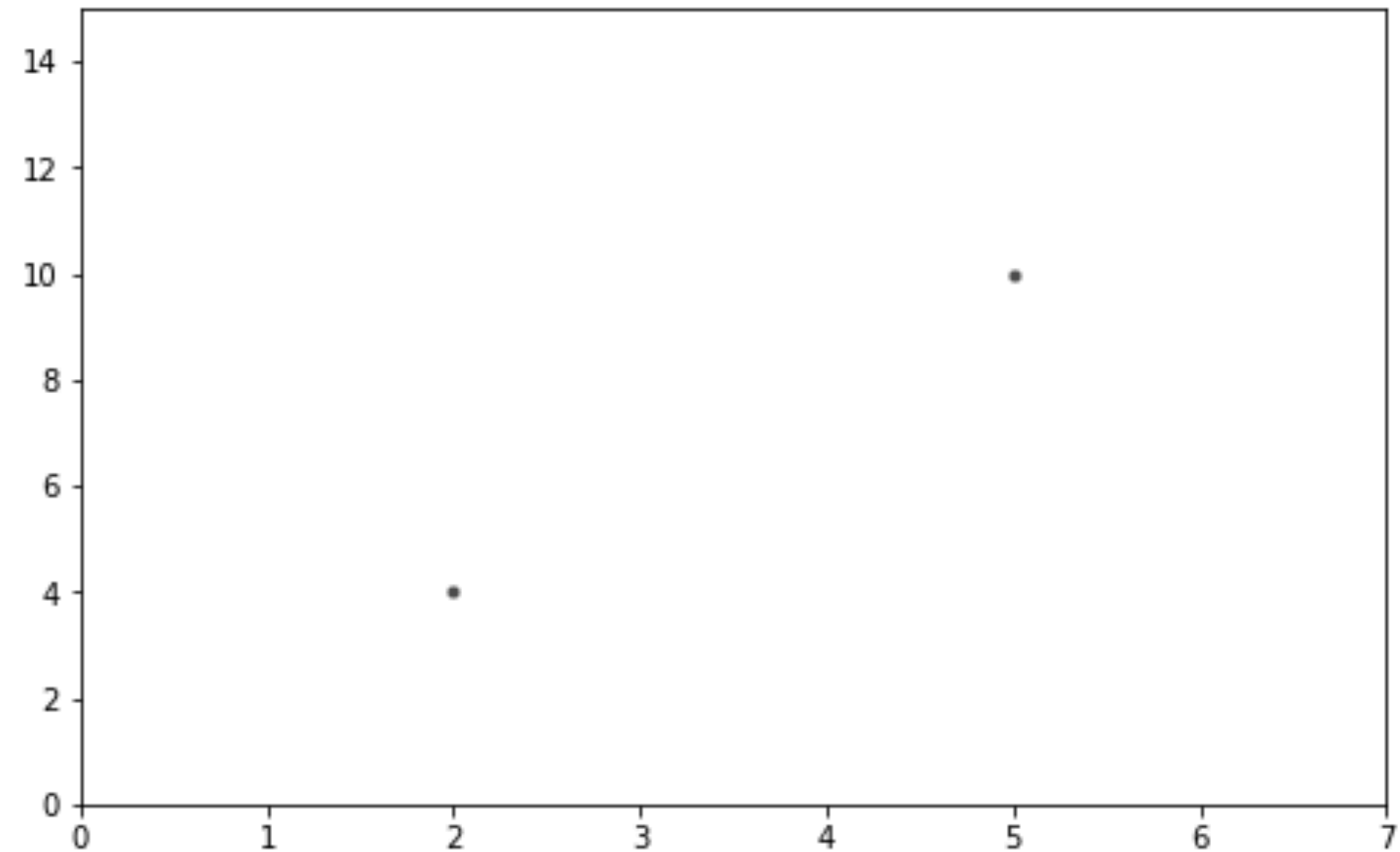
Galton le llamó a este fenómeno **regresión**, como en:

“La estatura del hijo tiende a ‘regresar’ (o desplazarse hacia) la estatura media (promedio)”



Ejemplo

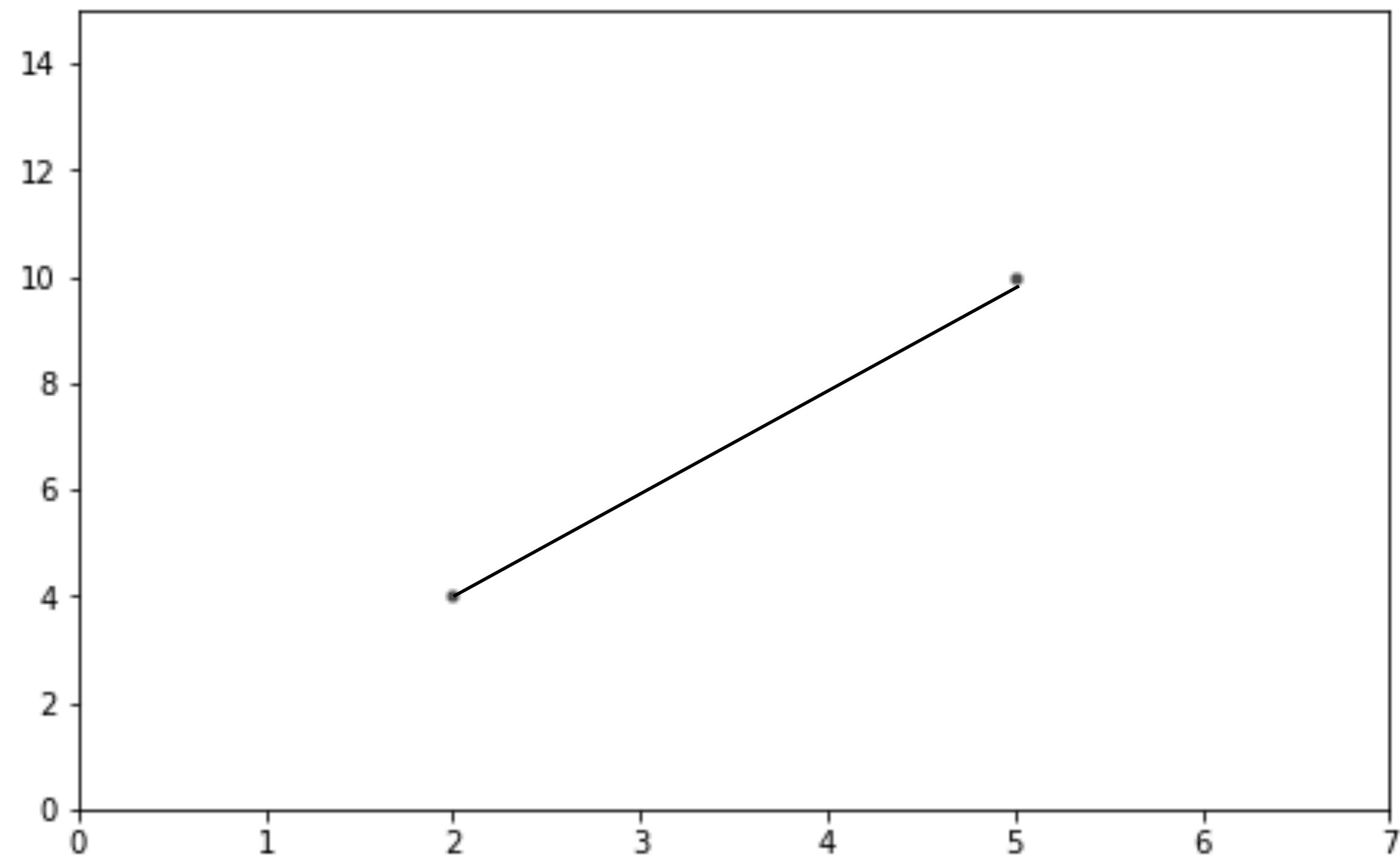
Usemos el ejemplo más simple posible: calcular la regresión con solo dos puntos.



Ejemplo

Lo único que estamos tratando de hacer es calcular nuestra línea de regresión de tal forma que pase tan cerca como sea posible a cada punto.

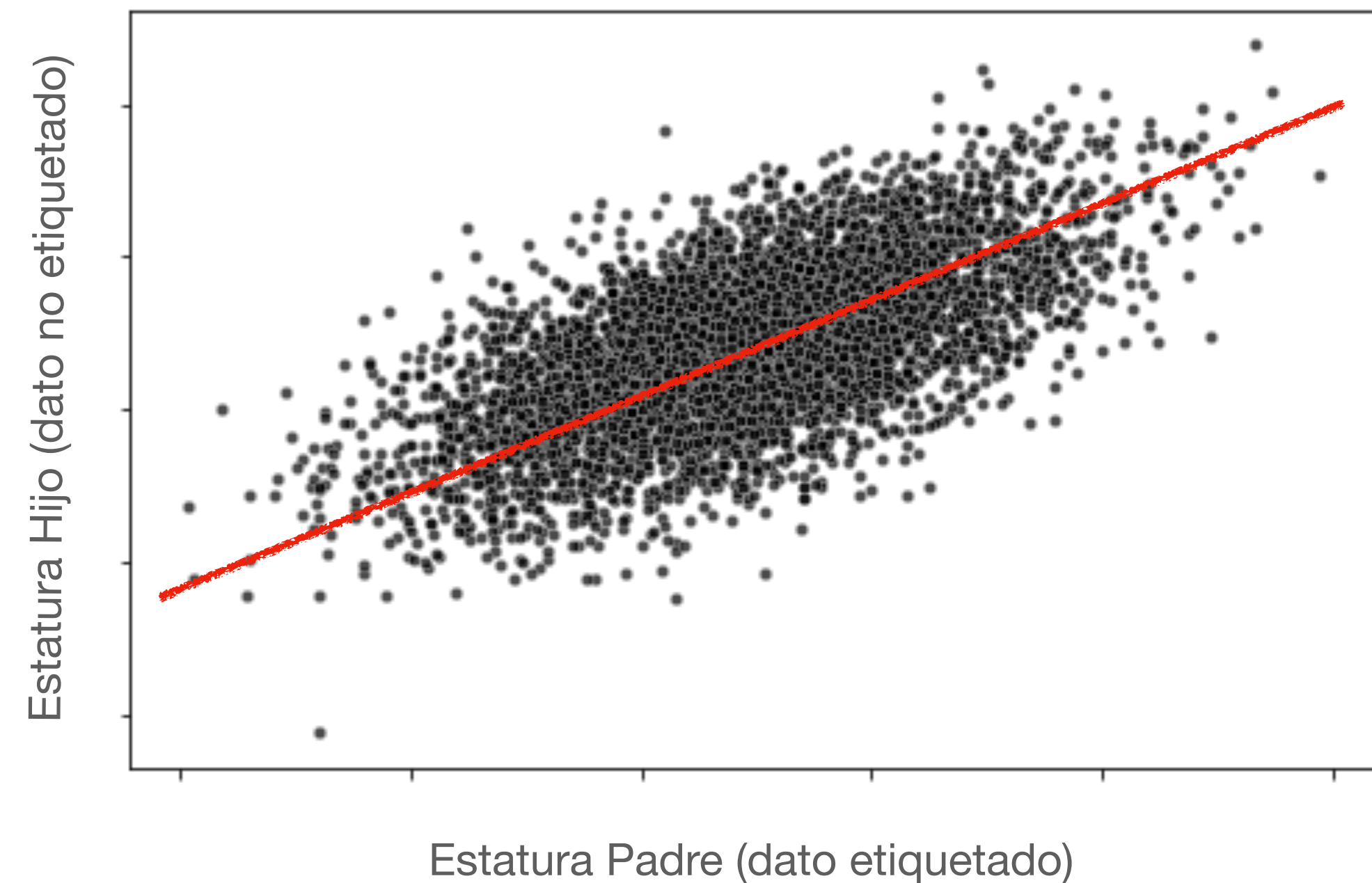
Para la regresión lineal clásica, o el Método de Mínimos Cuadrados (Least Squares Method), sólo se mide lo cercano de la línea a los puntos en una dirección vertical.



Ejemplo

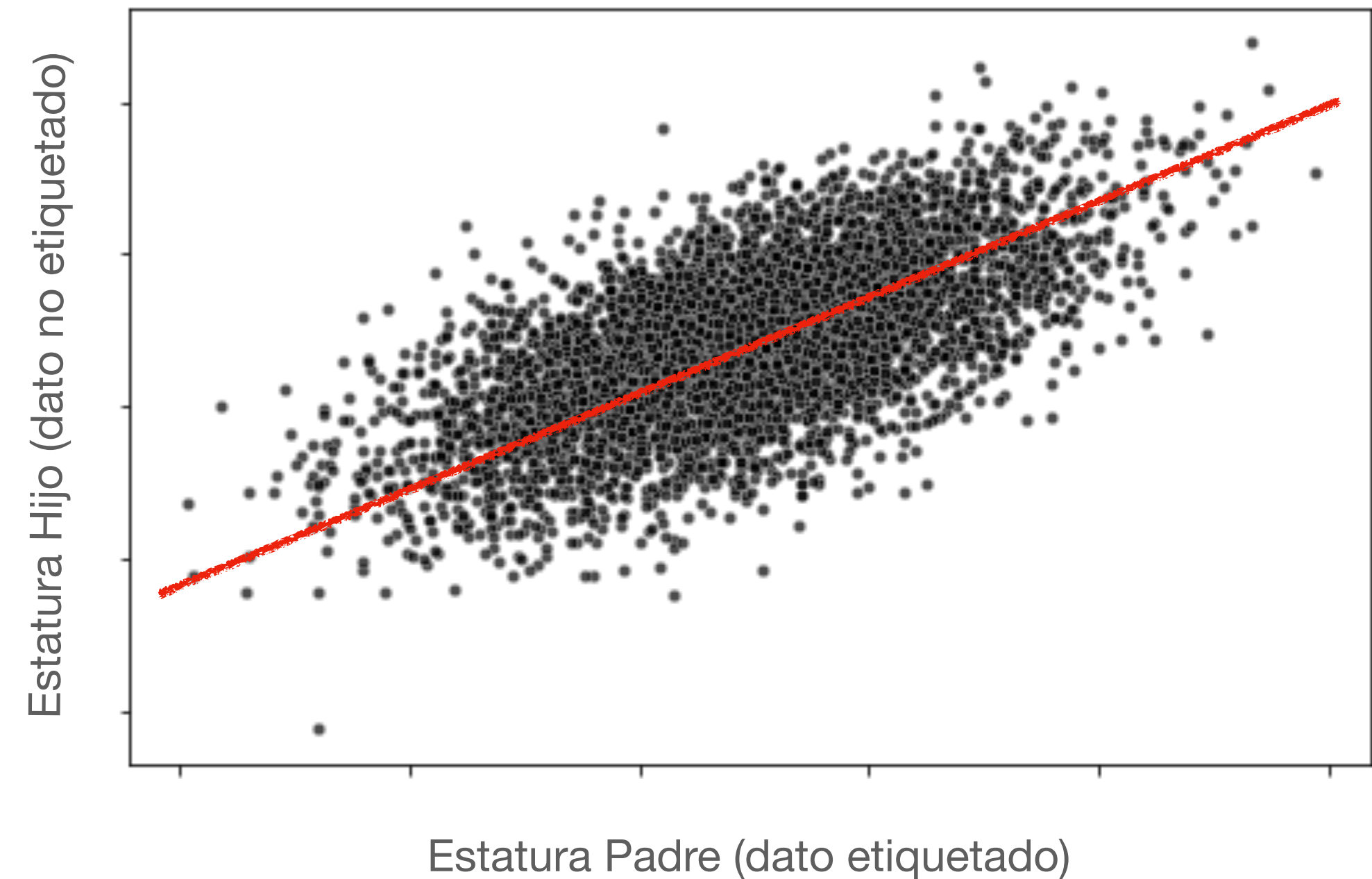
Si tan solo pudiéramos aplicar este mismo concepto a una gráfica con más de dos puntos.

Al hacer esto, podríamos tomar las estatura de muchos hombres y las de sus hijos y decirle a un hombre qué tan alto esperamos que sea su hijo...antes de que siquiera nazca!



Ejemplo

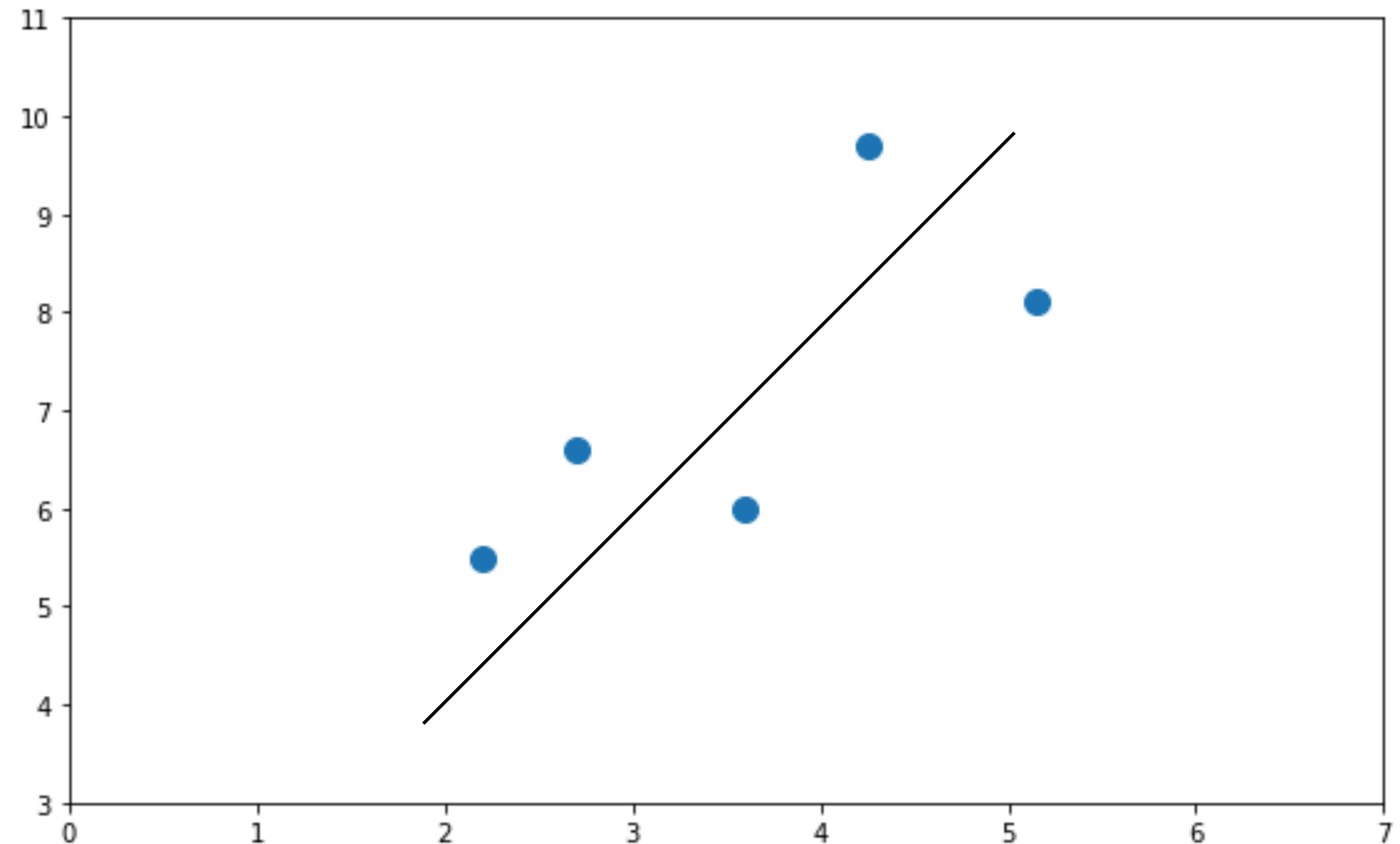
Hay varias formas de minimizar esto, (suma de errores cuadrados, suma de errores absolutos, etc.), pero todos estos métodos tienen una meta general de minimizar esta distancia.



Ejemplo

Ahora queremos ajustar (tallar) una línea de regresión lineal.

La pregunta es cómo decidir qué línea es la que mejor se ajusta?



Ejemplo

Utilizaremos el Método de Mínimos Cuadrados, que se talla o ajusta a través de minimizar la **suma de los cuadrados de los residuos**.

El residuo para una observación es la diferencia entre la observación misma (el valor y) y la línea ajustada

