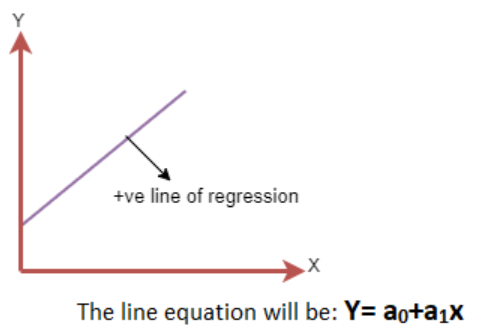
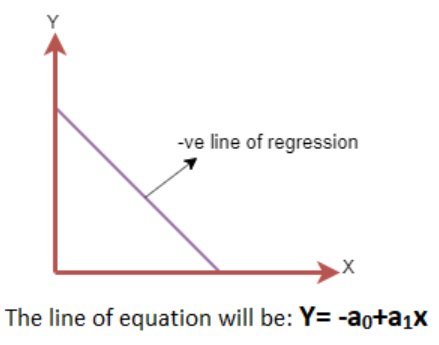
**LINEAR REGRESSIONS**

**Linear Regression Line**

* **Positive & negative regression:**

**Finding Best Fit Line**

* Our main goal is to find an **error free** line.

Cost function:-

* Used for estimating value of the **best fit** line.
* Or simply, ***cost function*** is used for checking **accuracy** of the **hypothesis function**.
* **Hypothesis function:** A function used for **mapping** **graph** (predicting).

**Error Handling**

* Mean squared error (MSE) is a type of cost function.
* It is calculated using **average of all squared errors**.
* **Residuals:** Distance between **predicted value** & **actual value**.
* Residuals directly depend on its distance from ***regression line***.

****

**N = Number of observations**

**Yi = Actual value**

**(a1xi + a0) = Predicted value**

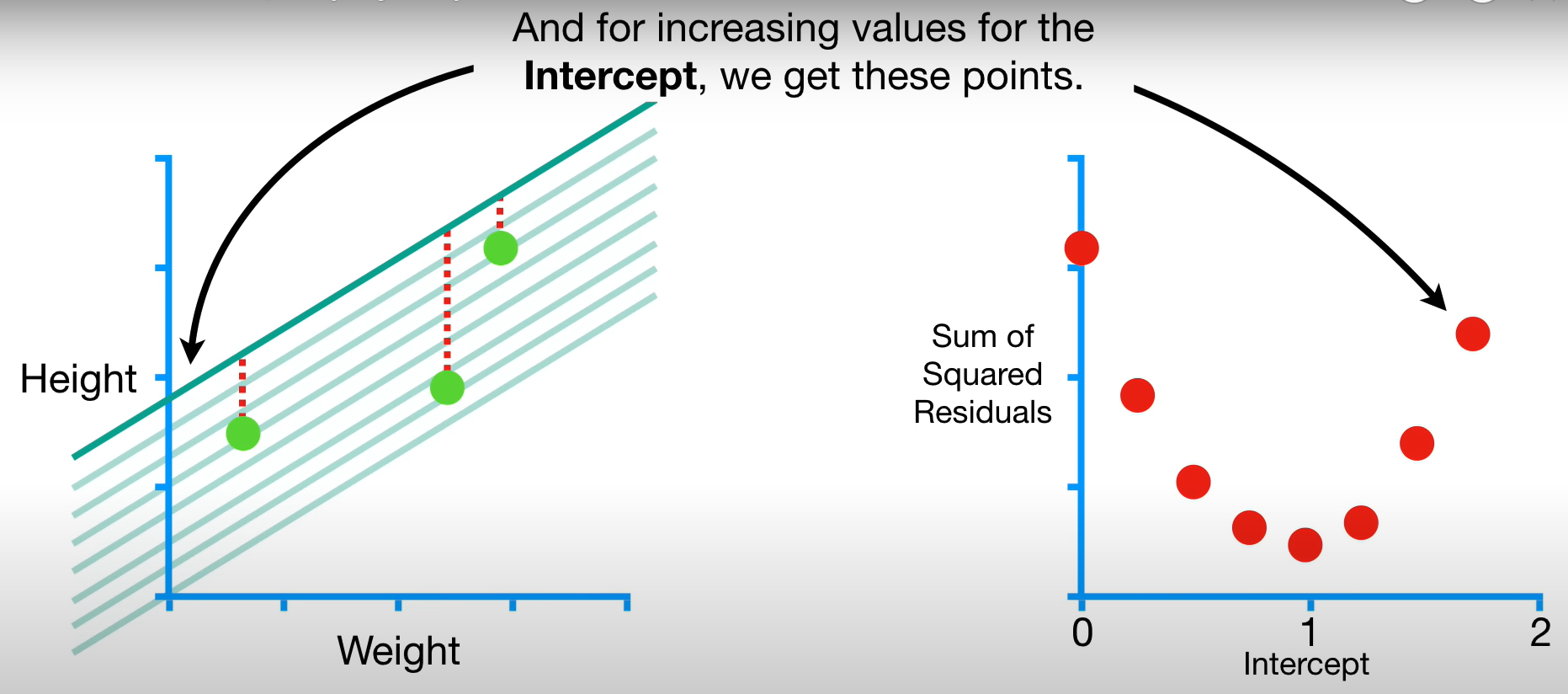
* This **sum of squared residuals (SSR)** provides us **accuracy** of the model.
* Formula of ***SSR*** is same as ***MSE***.
* The **less** the value, the **more** the accuracy.

**Gradient Descent**

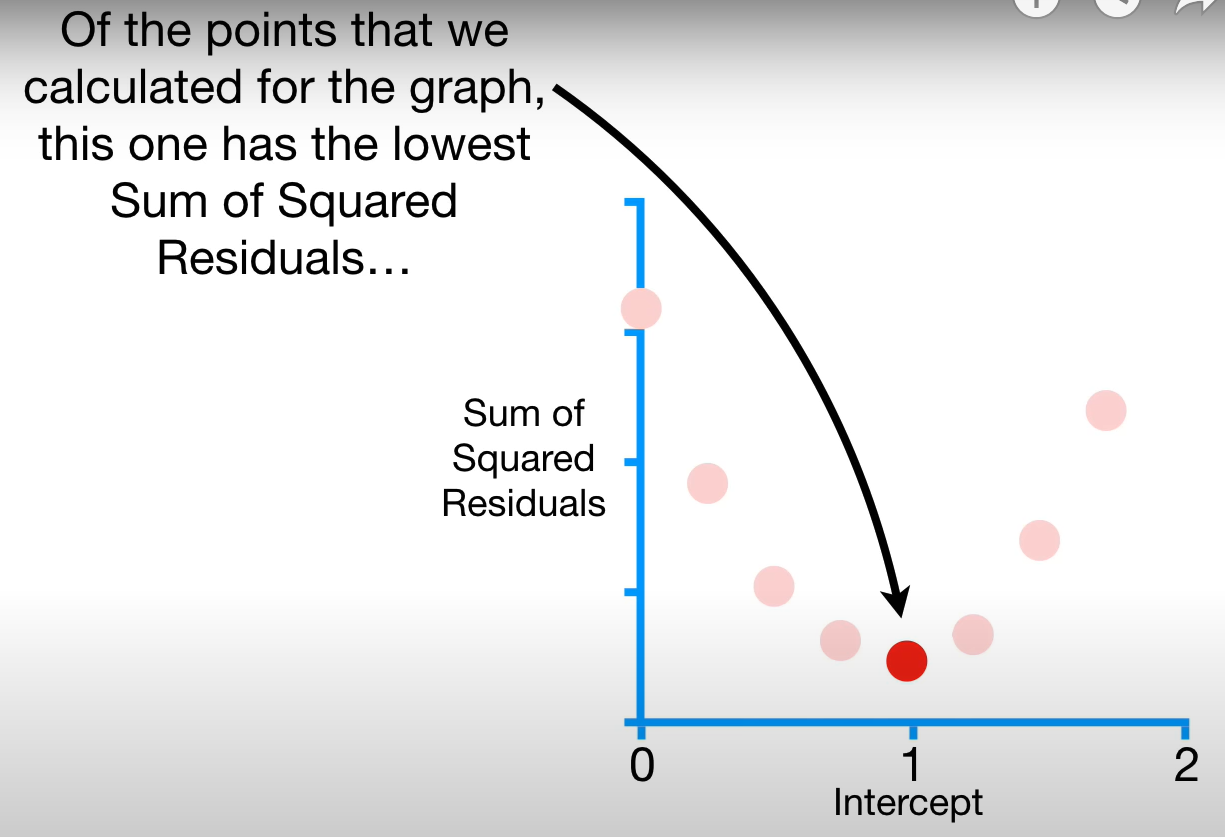
* It is used for **minimizing** value of **MSE** using gradient of ***cost function***.
* ***Gradient descent*** is used for **updating** coefficients of line by reducing ***cost function***.
* On graph, it provides us a **curve like** structure & we try to find the **least** SSR.

Steps for finding:-

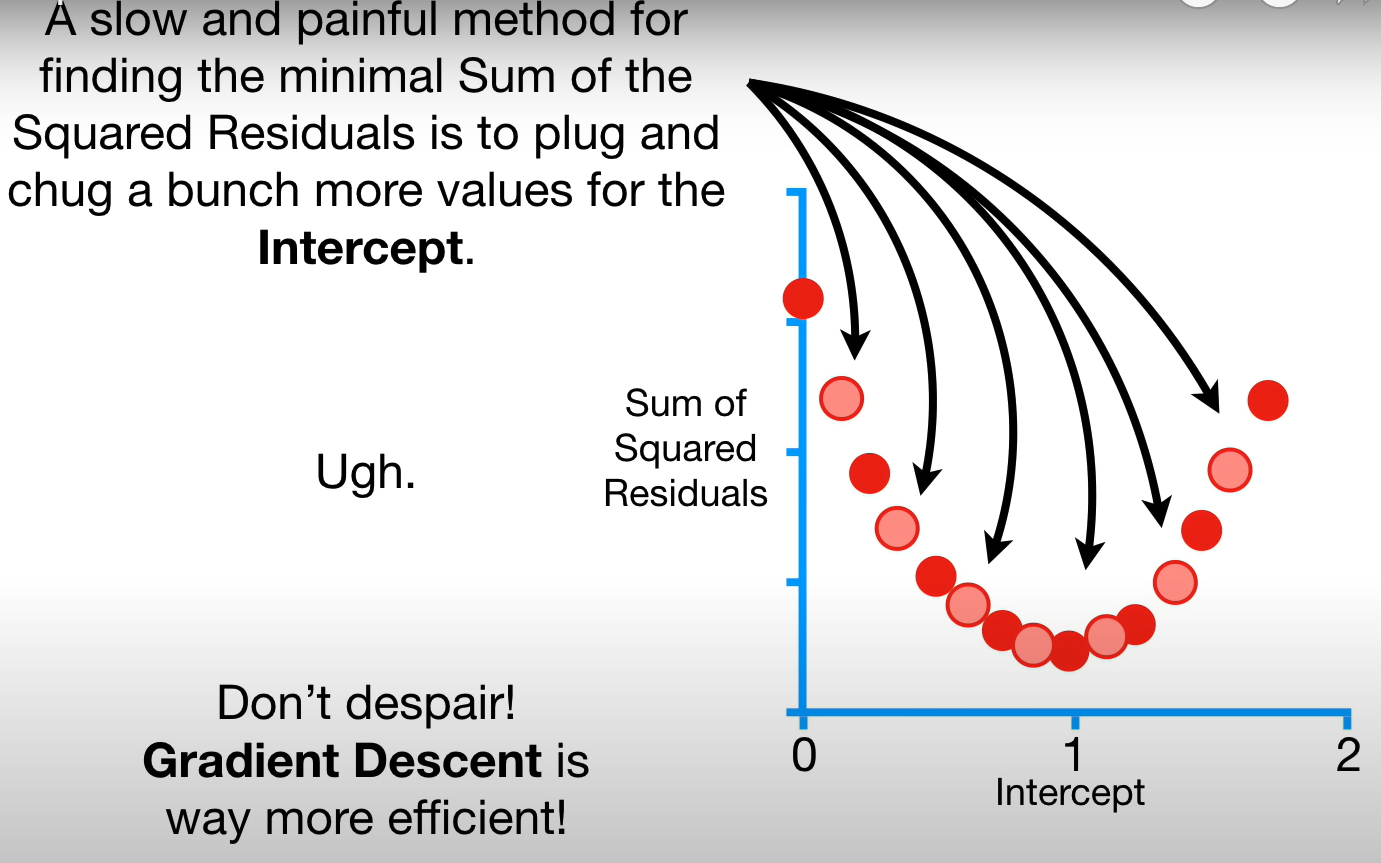
* **Step 1:** Plot the graph of all obtained SSRs.



* **Step 2:** Locate the **lowest** SSR value on graph.



* **Step 3:** Try plotting more SSRs between both **neighbour points** of the lowest SSR.



* **Step 4:** Pick the lowest SSR among the now available SSRs.

**Model Performance**

* **Optimization:** Process of finding best model out of all.
* It can be achieved through ***R-squared*** method.

**R-Squared Method**

* Tells **how well** the line fits with datapoints.
* It gives ratings for relationships between **dependent** & **independent variable** on a scale of **0-100%**.
* The **more** the value of ***R-square***, the **well** the line fits.
* Also known as ***coefficient of determination***.

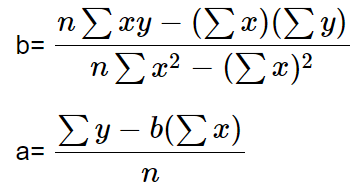
**R-squared = Explained variable / Total variation**

**Assumptions of Linear Regression**

***\*For getting best results from model, it’s advised to check these\****

* **Linear relationship** between dependent & independent variables.
* **Less or no multicollinearity** between dependent & independent variables.
* Assumes **homoscedasticity**, meaning assumes an error term to be applicable on all independent variables.
* **Normal distribution of errors** without much fluctuation in patterns.
* **No autocorrelation**, means no relation among error terms.

**Simple Linear Regression**



**y = a + bx + Ԑ**

**a = Intercept**

**b = Slope**

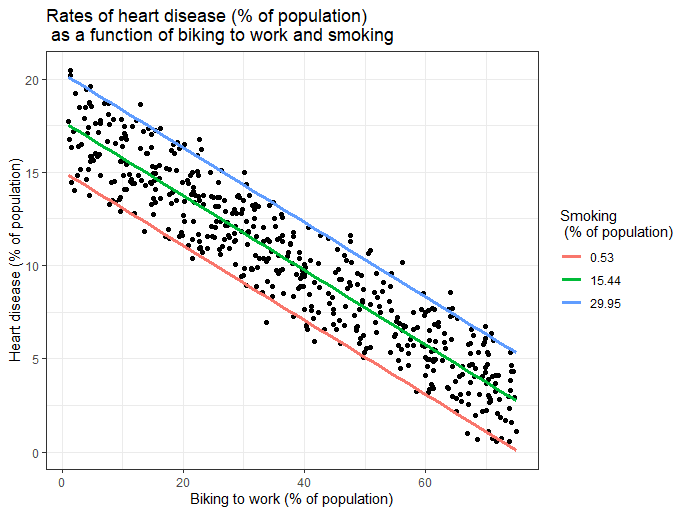
**Ԑ = Error term (negligible for good term)**

**Multiple Linear Regression (MLR)**

* In MLR, the predictor variables can be continuous or categorical.

**Equation:-**

**y = b0 + b1x + b2x + b3x + … + bnx**



**Assumptions of MLR**

* **Target** & **predictor** variables have linear relationship.
* **No collinearity** among independent variables.