Регрессионным анализом называют область математической статистики, связанную с выявлением и аналитическим выражением зависимостей между одной несколькими неслучайными величинами $(x_1, x_2, ..., x_n) = \mathbf{x}$ (*входными переменными*, *варьируемыми факторами* или *предикторами*) и доступной для измерения случайной величиной y (*откликом*).

Аналитическую зависимость $y = f(\mathbf{x})$ отклика от варьируемых факторов называют *экспериментально-статистической моделью*, *статистической моделью*, *регрессионной моделью* или *регрессией*. Общий вид этой зависимости $y = f(\mathbf{x}, b_1, b_2, ..., b_L)$ выбирается эвристически, на основе доступной информации об объекте исследования, и включает в качестве переменных не только предикторы $\mathbf{x}$, но и L неизвестных параметров b_1, b_2, …, b_L, подлежащих определению на основе экспериментальных данных $\{(\mathbf{x}, y)\}$. Включающую параметры модель называют *параметрической*.

Методологической основой нахождения параметров является *принцип максимального правдоподобия*: *наилучшим описанием системы является такое, при котором для модели максимальна вероятность предсказания отклика*. При предположениях о независимости, равной точности и нормальном распределении результатов измерений принцип максимального правдоподобия приводит к *методу наименьших квадратов*: *наилучшей моделью будет такая, для которой сумма квадратов отклонений эмпирических значений от значений, предсказанных моделью, обращается в минимум*.

Пусть в процессе исследования варьировалась одна независимая переменная x и после проведения N экспериментов получены значения y_u, $u = \overline{1, N}$. Требуется методом наименьших квадратов подобрать параметры линейной регрессионной модели

$$y = ax + b.$$

Сумма квадратов отклонений эмпирических и предсказанных значений:

$$S = \sum_{u=1}^{N} \left(y_u - (ax_u + b) \right)^2.$$

Считая эту сумму функцией неизвестных параметров $S = S(a,b)$, потребуем выполнения необходимого условия локального экстремума:

$$\begin{cases} \dfrac{\partial S}{\partial a} = 0 \\ \dfrac{\partial S}{\partial b} = 0 \end{cases}.$$

Дифференцируя и приравнивая частные производные к нулю, получим:

$$\begin{cases} \sum\limits_{u=1}^{N} \left(y_u - a x_u - b \right) x_u = 0 \\ \sum\limits_{u=1}^{N} \left(y_u - a x_u - b \right) = 0 \end{cases}.$$

Изменим порядок суммирования:

$$\begin{cases} Nb + \left(\sum\limits_{u=1}^{N} x_u \right) a = \sum\limits_{u=1}^{N} y_u \\ \left(\sum\limits_{u=1}^{N} x_u \right) b + \left(\sum\limits_{u=1}^{N} x_u^2 \right) a = \sum\limits_{u=1}^{N} y_u x_u \end{cases}.$$

Если все значения x_u различны, то полученная система двух линейных уравнений (которую называют *нормальной системой*) имеет единственное решение. Можно доказать, что это решение действительно соответствует точке локального минимума функции $S = S(a,b)$.