

Линейные регрессионные модели

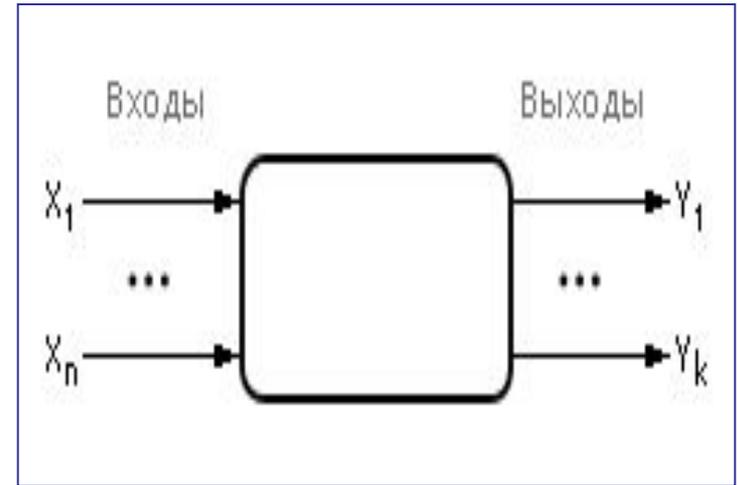
Одномерная регрессионная модель
Многомерная регрессионная модель

По степени информированности исследователя об объекте существует деление объектов на три типа «ящиков»:

«белый ящик»: об объекте известно все;

«серый ящик»: известна структура объекта, неизвестны количественные значения параметров;

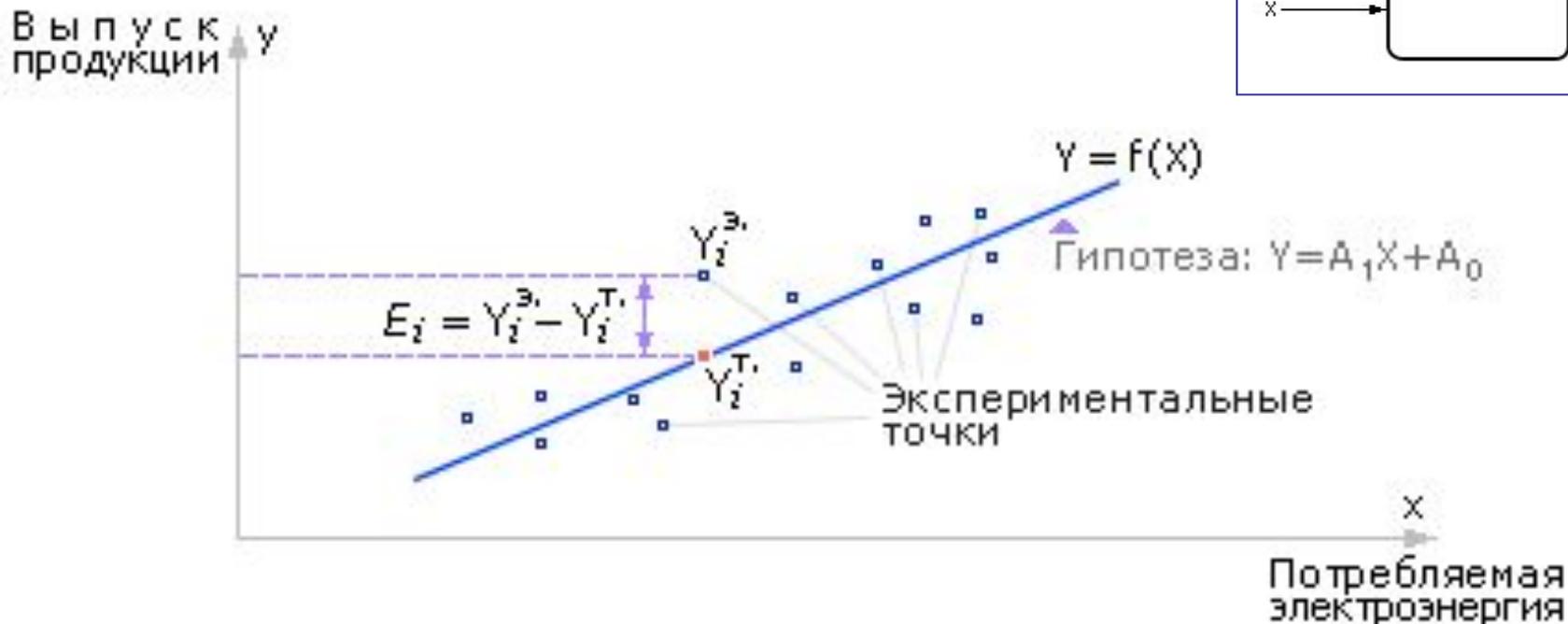
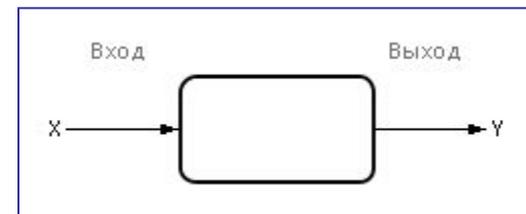
«черный ящик»: об объекте неизвестно ничего.



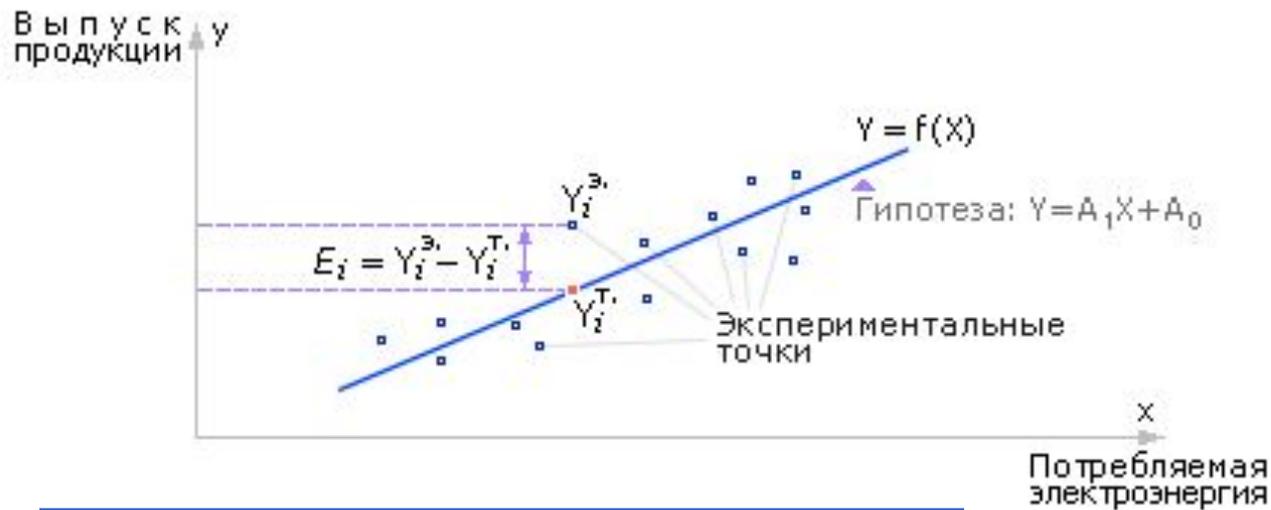
Обозначение черного ящика на схемах

Задача состоит в том, чтобы, зная множество значений на **входах и выходах**, построить **модель**, то есть определить функцию ящика, по которой вход преобразуется в выход. Такая задача называется **задачей регрессионного анализа**.

Пусть, например, перед нами стоит задача определить, как зависит выпуск продукции от количества потребляемой электроэнергии. Результаты наблюдений отобразим на графике/ Всего на графике n экспериментальных точек, которые соответствуют n наблюдениям



Для начала предположим, что мы имеем дело с черным ящиком, имеющим один вход и один выход. Допустим для простоты, что зависимость между входом и выходом линейная или почти линейная. Тогда данная модель будет называться **линейной одномерной регрессионной моделью**.



Рассматривая экспериментально полученные данные, **предположим**, что они подчиняются **линейной гипотезе**, то есть выход Y зависит от входа X линейно, то есть гипотеза имеет вид:

$$Y = A_1X + A_0$$

$$A_0 = \frac{\sum_{i=1}^n Y_i \sum_{i=1}^n X_i^2 - \sum_{i=1}^n X_i Y_i \sum_{i=1}^n X_i}{n \sum_{i=1}^n X_i^2 - (\sum_{i=1}^n X_i)^2}$$

$$A_1 = \frac{n \sum_{i=1}^n X_i Y_i - \sum_{i=1}^n Y_i \sum_{i=1}^n X_i}{n \sum_{i=1}^n X_i^2 - (\sum_{i=1}^n X_i)^2}$$

Проверка линейной гипотезы

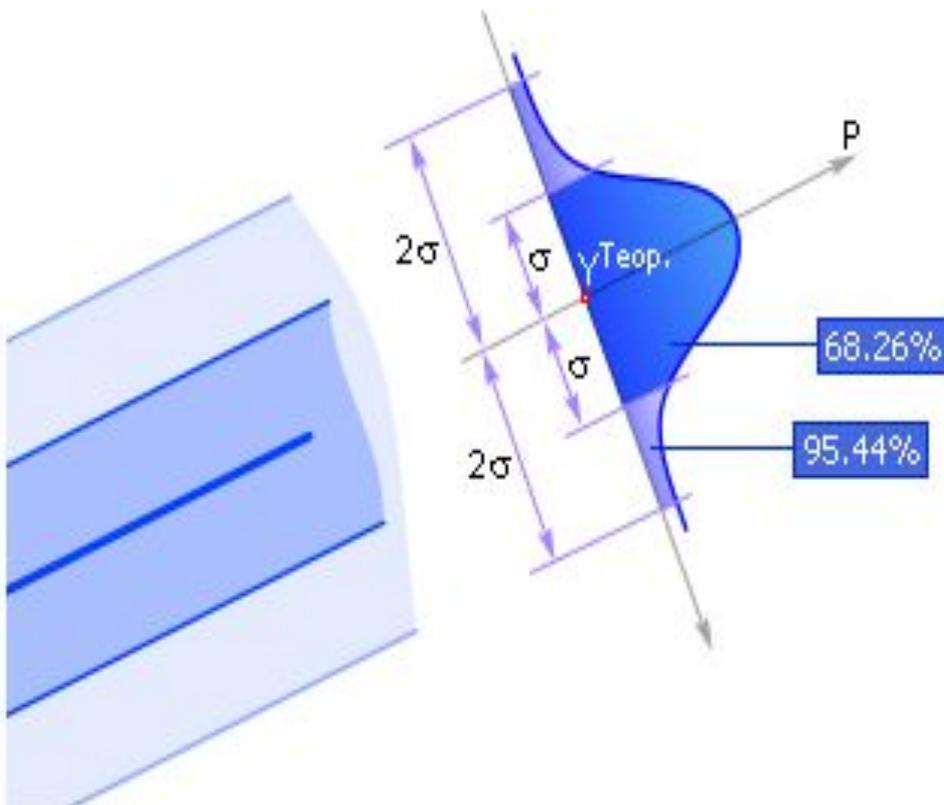
Чтобы определить, принимается гипотеза или нет, нужно, во-первых, рассчитать ошибку между точками заданной экспериментальной и полученной теоретической зависимости и суммарную ошибку:

$$E_i = (Y_i^{\text{Эксп.}} - Y_i^{\text{Теор.}}), i = 1, \dots, n$$

$$F(A_0, A_1) = \sum_{i=1}^n E_i^2$$

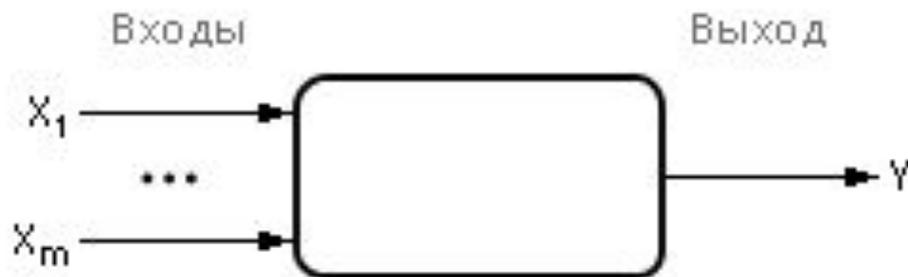
$$\sigma = \sqrt{\frac{F}{n}}$$

n-общее
число точек



Если в полосу, ограниченную 2σ попадает 68.26% и более экспериментальных точек то выдвинутая гипотеза принимается. В противном случае выбирают более сложную гипотезу или проверяют исходные данные. Если требуется большая уверенность в результате, то используют дополнительное условие: в полосу, ограниченную линиями 4σ , должны попасть 95.44% и более экспериментальных точек.

Линейная множественная модель



Гипотеза – линейная модель

$$Y = A_0 + A_1 \cdot X_1 + \dots + A_m \cdot X_m$$

Для нахождения коэффициентов A_i методом Крамера представим систему в матричном виде:

$$\begin{pmatrix}
 \sum_{i=1}^n X_{1i} & \sum_{i=1}^n X_{2i} & \dots & \sum_{i=1}^n X_{mi} \\
 \sum_{i=1}^n X_{1i} \sum_{i=1}^n X_{1i} & \sum_{i=1}^n X_{2i} \sum_{i=1}^n X_{1i} & \dots & \sum_{i=1}^n X_{mi} \sum_{i=1}^n X_{1i} \\
 \sum_{i=1}^n X_{2i} \sum_{i=1}^n X_{1i} & \sum_{i=1}^n X_{2i} \sum_{i=1}^n X_{2i} & \dots & \sum_{i=1}^n X_{mi} \sum_{i=1}^n X_{2i} \\
 \dots & \dots & \dots & \dots \\
 \sum_{i=1}^n X_{mi} \sum_{i=1}^n X_{1i} & \sum_{i=1}^n X_{2i} \sum_{i=1}^n X_{mi} & \dots & \sum_{i=1}^n X_{mi} \sum_{i=1}^n X_{mi}
 \end{pmatrix}
 \cdot
 \begin{pmatrix}
 A_0 \\
 A_1 \\
 A_2 \\
 \dots \\
 A_m
 \end{pmatrix}
 =
 \begin{pmatrix}
 \sum_{i=1}^n Y_i \\
 \sum_{i=1}^n Y_i X_{1i} \\
 \sum_{i=1}^n Y_i X_{2i} \\
 \dots \\
 \sum_{i=1}^n Y_i X_{mi}
 \end{pmatrix}$$

Конец темы