

Значимость регрессионной модели (Significance of regression model)

Значимостью регрессионной модели называют степень статистической связи между ее входной и выходной переменными. Если такая связь полностью отсутствует, то поведение входной переменной никак не «объясняет» поведение выходной, и модель оказывается бесполезной.

Для простой линейной регрессии это означает, что в уравнении $y = b + ax$ коэффициент $a = 0$ и линия регрессии будет параллельна оси x .

Такая модель для любых значений входной переменной x всегда будет выдавать одну и ту же оценку выходной переменной y , равную ее среднему значению, вычисленную по всем наблюдениям. Очевидно, что такая модель будет, скорее всего, бесполезной. С таким же успехом для оценки выходной переменной можно использовать ее среднее значение.

Если коэффициент a отличен от 0, то линия регрессии имеет наклон, что указывает на зависимость выходной переменной от входной, а сама линия аппроксимирует эту зависимость. В этом случае модель будет иметь некоторую значимость. С целью оценки ее уровня для регрессионной модели используются коэффициент детерминации и коэффициент корреляции.

На практике значимость регрессионной модели определяется значимостью ее коэффициентов, т.е. требуется провести тестирование с целью выяснить, значимо ли ее коэффициент отличается от нуля для конкретного набора данных.

Для этого используются специальные статистические таблицы, в которых для заданного числа наблюдений выборки n и числа параметров модели k , а точнее числа степеней свободы $(n - k)$, определяется критическое значение статистики и сравнивается с уровнем значимости (например, 5%).

Если критическое значение меньше уровня значимости, то нулевую гипотезу о незначимости коэффициента регрессии следует отвергнуть и принять альтернативную, т.е. коэффициент значим.