

Тест Чоу (Chow test)

Раздел: [Метрики](#)

Статистический тест, позволяющий оценить значимость улучшения регрессионной модели после разделения исходной выборки на части. Одним из ограничений линейной регрессии является то, что для различных интервалов значений независимой переменной характер ее связи с выходной переменной может меняться.

Например, с увеличением возраста клиента его кредитный рейтинг может возрастать, но до определенного порога (например 50—55 лет), после которого люди начинают чаще болеть, им сложнее найти работу и т.д. Они становятся менее привлекательны для банка в качестве заемщиков, и для них имеет место обратная тенденция.

Очевидно, что любая модель, которая аппроксимирует такую зависимость единственной линейной функцией, вряд ли будет точной. Выходом из ситуации является разделение диапазона значений независимой переменной на два, в пределах каждого из которых зависимость между ней и выходной переменной монотонна. Затем для каждого из полученных интервалов строится свое уравнение регрессии.

Возникает вопрос: как разделить исходное множество так, чтобы обеспечить лучшую аппроксимацию? Для этого обычно строят множество разбиений, для каждого определяют значимость улучшения модели и выбирают то из них, для которого она максимальна. Для ее оценки и используется тест Чоу, который применяет статистику:

$$\frac{(S - S_1 - S_2)/(k+1)}{(S_1 + S_2)/(n - 2k - 2)},$$

где S — сумма квадратов остатков линейной регрессии целой модели, S_1 — сумма квадратов остатков модели, построенной для первого подмножества, S_2 — сумма квадратов остатков модели, построенной для второго подмножества, n — общее число наблюдений выборки.

Полученная статистика подчиняется F — распределению Фишера с $k + 1$ и $n - 2k + 1$ степенями свободы и используется для оценки значимости улучшения модели при разделении выборки на две части.