

## ОСНОВНЫЕ ПОКАЗАТЕЛИ КАЧЕСТВА МОДЕЛЕЙ РЕГРЕССИИ

Нессонова М.Н.

*Национальный фармацевтический университет,  
г. Харьков*

Модели регрессии широко используются в современных приложениях и информационных технологиях поддержки принятия решений в различных областях. Применение данного метода моделирования подразумевает наличие набора прецедентов, в котором целевой показатель измерен в интервальной шкале, а основной задачей является вывод решающего правила, позволяющего предсказывать его значения на основании значений объясняющих признаков. Результатом решения подобных задач обычно является несколько возможных моделей одного процесса, полученных разными способами, и, следовательно, немаловажной проблемой есть выбор наилучшей из них. К основным показателям, на основании которых осуществляется этот выбор, относятся коэффициент детерминации ( $R^2$ ), средняя абсолютная и среднеквадратическая ошибки, и среднее абсолютное отклонение. Значение  $R^2$  трактуется как доля изменчивости целевого показателя, объяснённая построенной моделью. Часто его переводят в проценты и говорят, что он показывает, на сколько процентов модель соответствует исходным данным. Значения средних абсолютной и квадратичной ошибок используют для характеристики на сколько единиц в среднем отклоняется модель от исходных данных. Средняя относительная ошибка показывает это отклонение в процентах относительно величины целевого показателя. Благодаря понятному физическому смыслу эти показатели являются наиболее часто используемыми для оценки качества регрессионных моделей, тем не менее, их применение имеет свои особенности, знать которые и учитывать при разработке практических приложений будет полезным как для начинающих, так и для опытных разработчиков математического обеспечения информационных и компьютерных систем прогнозирования и поддержки принятия решений. Так, следует иметь в виду, что оценка качества регрессионной модели по значению  $R^2$  сильно обусловлена предметной областью и самой прикладной задачей, для которой разрабатывается модель. Кроме того, рекомендуют избегать ситуаций, когда  $R^2$  приближается к 1, т.к. подобные модели часто противоречат принципу регуляризации и являются "переобученными" в силу чего оказываются непригодными для прогнозирования новых значений. От использования средней относительной ошибки следует отказаться в случаях, когда реальные значения целевой переменной содержат нулевые и близкие к нулю значения, т.к. она может оказываться неадекватно большой. Для оценки ошибки регрессии в абсолютных единицах следует предпочесть среднюю квадратичную ошибку, если важно исключить модели, имеющие экстремальные отклонения; если же наличие единичных больших погрешностей не является критичным, лучше использовать среднюю абсолютную ошибку.