

Е. А. Драенкова, Ю. В. Синюгина
(УО «ГГУ им. Ф. Скорины», Гомель)

ЛОГИСТИЧЕСКАЯ РЕГРЕССИЯ И РОС-АНАЛИЗ В СРЕДЕ R

На сегодняшний день статистическая среда R является лидером среди некоммерческих систем статистического анализа и постепенно становится незаменимой при проведении научных и технических расчетов в большинстве западных университетов и многих ведущих фирмах. Благодаря большому разнообразию реализованных в R алгоритмов машинного обучения ее все чаще выбирают для построения предсказательных моделей.

В R для построения модели логистической регрессии – статистической модели для предсказания вероятности возникновения интересующего события – используется функция `glm()`:

```
model <- glm(dependent_variable ~ predictor1+ predictor2+ predic-  
tor3+..., family=binomial(logit), data)
```

Совместно с функцией `glm()` могут использоваться следующие функции:

`summary(model)` – вывод информации о построенной модели;

`model.reduced <- step(model)` – построение редуцированной модели;

`anova(model.reduced, model, test="Chisq")` – сравнение редуциро-
ванной и полной моделей;

`coef(model)` – вывод регрессионных коэффициентов;

`exp(coef(model))` – потенцирование регрессионных коэффициентов;

`predict(model, newdata=testdata, type="response")` – построение про-
гноза модели.

Неотъемлемой частью логистической регрессии является ROC – анализ – оценка прогностической способности модели. Перед запуском ROC – анализа необходимо подключить пакет `rROC`. Для расчета и построения ROC – кривой модели используются функции `roc()` и `plot.roc()` соответственно:

```
roc.model <- roc( dependent_variable, predict(model,  
type="response"), ci=TRUE)
```

```
plot.roc(roc.model, print.auc=TRUE, print.thres=TRUE)
```

Сравнение построенных моделей осуществляется с помощью непараметрического теста Делонга командой `roc.test(roc.model1, roc.model2, method="delong")`.