

2019
Том 2

Министерство науки и высшего образования
Российской Федерации

ФГБОУ ВО «Владивостокский государственный
университет экономики и сервиса»



17–19 апреля
2019 г.

ИНТЕЛЛЕКТУАЛЬНЫЙ ПОТЕНЦИАЛ ВУЗОВ –

НА РАЗВИТИЕ ДАЛЬНЕВОСТОЧНОГО РЕГИОНА РОССИИ И СТРАН АТР

Материалы XXI Международной
научно-практической
конференции студентов,
аспирантов и молодых ученых

В пяти томах

Том 2



ИНТЕЛЛЕКТУАЛЬНЫЙ ПОТЕНЦИАЛ ВУЗОВ – НА РАЗВИТИЕ ДАЛЬНЕВОСТОЧНОГО РЕГИОНА РОССИИ И СТРАН АТР

giographic study is traumatic, the risk of complications is present. The work is devoted to the construction of a model of logistic regression, which can be an alternative to coronary angiography in the diagnosis of coronary heart disease.

Keywords: coronary heart disease, logistic regression, ROC curve, logit model.

В настоящее время сердечно-сосудистые заболевания являются основной причиной смертности и инвалидности во всем мире. Ведущая роль в структуре смертности от сердечно-сосудистых заболеваний принадлежит ишемической болезни сердца.

Ишемическая болезнь сердца (ИБС) – хроническое заболевание, которое развивается при недостаточном поступлении кислорода к миокарду. Основной причиной (более чем в 90% случаев) недостаточного поступления кислорода является образование атеросклеротических бляшек в просвете коронарных артерий, артерий кровоснабжающих сердечную мышцу (миокард) [1].

Коронарография остается «золотым стандартом» в диагностике ИБС, так как на сегодняшний день это единственный из существующих методов исследований, позволяющий определить точные детали анатомического строения всего коронарного русла.

Контрастирование коронарных сосудов — это наиболее достоверный способ для выбора тактики лечения пациентов с ишемией миокарда. Осложнения при этой процедуре хоть и достаточно редко, но встречаются. Диагностика связана с введением катетера в сосуды сердца, подачей через него контрастного вещества, поэтому она может представлять потенциальную опасность для больного. Для предотвращения нежелательных последствий нужно тщательное обследование и подготовка.

Так как диагностика коронарного кровотока подразумевает прокол периферической артерии бедра или плеча, заведение через него катетера, продвижение его по аорте и венечным сосудам, подачу контрастного йодсодержащего вещества, то это может сопровождаться негативной реакцией организма. Риск осложнений повышается, если больной страдает: артериальной гипертензией; сахарным диабетом; распространенным атеросклерозом; склонностью к аллергическим реакциям; патологией почек; нарушением ритма; ожирением или дефицитом массы тела; хроническим алкоголизмом; инфекционным заболеванием и т.д.

В зависимости от этапа коронарографии она может вызывать такие осложнения как: гематома; тромбоз; инфекция; аллергия; анафилаксия; интоксикация; поражение почек; падение свертывающей способности крови и, как следствие, кровотечение; аритмия; инфаркт; инсульт. Всё это может привести даже к летальному исходу[3]. Цель данного исследования заключается в построении модели логистической регрессии, позволяющей ставить диагноз ИБС (ишемическая болезнь сердца) без коронарографии. Для достижения поставленной цели необходимо выполнить следующие задачи:

1. Анализ и первичная обработка предоставленных данных;
2. Выбор предикторов, используемых в модели;
3. Построение модели логистической регрессии со значимыми переменными.

Логистическая регрессия – это статистическая модель, используемая для прогнозирования вероятности возникновения некоторого события путём подгонки данных к логистической кривой.

Модель логистической регрессии очень часто используется в медицине при проведении клинических исследований, в банковском скоринге и других прикладных задачах.

Логистическая регрессия применяется для прогнозирования вероятности возникновения некоторого события по значениям множества признаков. Для этого вводится так называемая *зависимая переменная* y , принимающая лишь одно из двух значений — как правило, это числа 0 (событие не произошло) и 1 (событие произошло), и множество *независимых переменных* (предикторов) — вещественных X_1, X_2, \dots, X_n , на основе значений которых требуется вычислить вероятность принятия того или иного значения зависимой переменной.

Для оценки вероятности $P(y = 1) = F(u)$ используем функцию логистического распределения $F(u) = \Lambda(u) = \frac{e^u}{1 + e^u}$ [3], где u – линейная комбинация предикторов.

Для построения модели использовались данные более 300 пациентов. В качестве зависимой переменной в данной работе является результат диагностирования ИБС методом коронарографии ($y = 0$, если поставлен диагноз ИБС, $y = 1$, в противном случае). Предикторами выступают различные факторы медицинского обследования и индивидуальных характеристик пациен-

тов. Объясняющие переменные в этой модели могут быть как категориальные, так и количественные. Все расчеты и построения графиков выполнены с помощью программы RStudio[4].

Предварительно данные были разделены на обучающую и тестовую выборку. В ходе работы решалась проблема пропущенных данных методом удаления соответствующих наблюдений. В результате построения логит-модели с различными множествами предикторов на данном этапе исследования получена мо-

$$\text{дель } \ln\left(\frac{P(y=1)}{P(y=0)}\right) = 0.61910 - 0.00058x_1 - 0.01623x_2 + 0.02686x_3,$$

где x_1 - фибриноген, г/л, x_2 - AST, x_3 - ALT. Модель со значимыми коэффициентами при всех переменных.

```
call:
glm(formula = "коронарোগен: без гемодинамически значимых стенозов (если 0 ИКС есть)" ~
"фибриноген, тр/л" + AST + ALT, family = binomial(link = "logit"),
data = d_tr216)

variance-covariance of fixed effects:
             (Intercept)             фибриноген, тр/л             AST             ALT
             0.61910000  -0.00058000 -0.01623000  0.02686000

variance-covariance of fixed effects:
             (Intercept)             фибриноген, тр/л             AST             ALT
             0.61910000  -0.00058000 -0.01623000  0.02686000

Coefficients:
             (Intercept)             фибриноген, тр/л             AST             ALT
             0.61910000  -0.0005790  -0.0162295  -0.298  0.000972 ***
             0.0162289  0.0073342  -2.345  0.024891 *
             0.0268649  0.0064021  2.830  0.001631 **

---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for binomial family taken to be 1)

null deviance: 283.82  on 233  degrees of freedom
Residual deviance: 263.80  on 230  degrees of freedom
(1 observation deleted due to missingness)
AIC: 269.8

number of Fisher scoring iterations: 4
```

Рис. 1. Результаты моделирования

Выводы о качестве модели также позволяет сделать ROC-анализ (Receiver Operating Characteristic) — это анализ прогностической эффективности модели.

По ROC-кривой делают вывод на основании площади под кривой (чем больше площадь, тем лучше). ROC-кривая — это график, позволяющий оценить качество бинарной классификации, отображает соотношение между долей объектов от общего количества носителей признака, верно классифицированных как несущих признак (TPR, называемой чувствительностью алгоритма классификации) и долей объектов от общего количества объектов, не несущих признака, ошибочно классифицированных как несущих признак. (величина 1-FPR называется специфичностью алгоритма классификации) при варьировании порога решающего правила.

Количественную интерпретацию ROC даёт показатель AUC — площадь, ограниченная ROC-кривой и осью доли ложных положительных классификаций. Чем выше показатель AUC, тем качественнее классификатор.

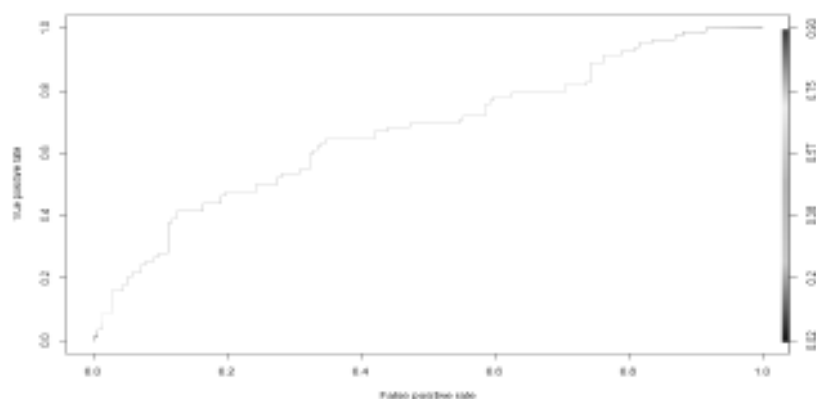


Рис. 2. ROC-кривая

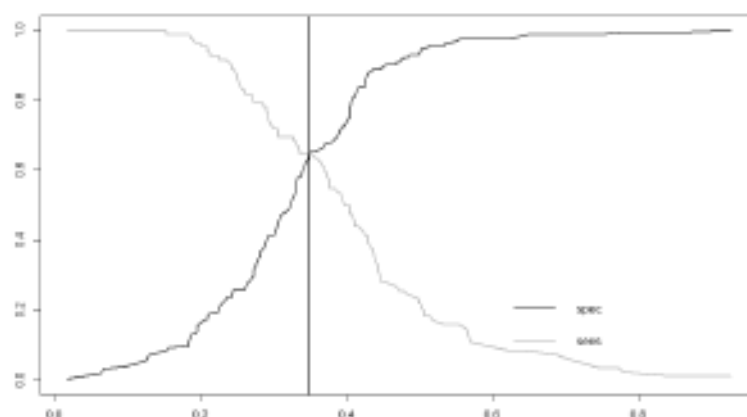


Рис. 3. Порог классификации

По сопоставлению кривых чувствительности и специфичности оптимальный порог классификации может быть выбран примерно 0.38

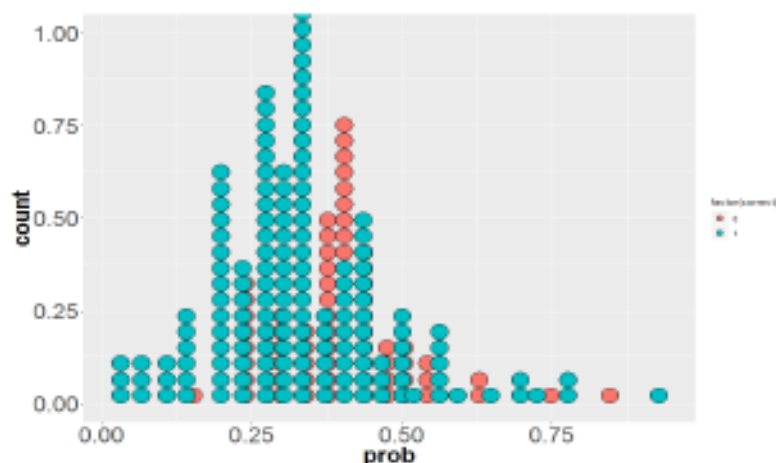


Рис.4. Итоги

На данный момент нами получены следующие результаты.

Процент правильно классифицированных моделью наблюдений равен 0.6367521 (63%) на обучающей выборке и 0.5921053 (59%) на тестовой.

Данные постоянно обновляются и добавляются, это позволяет нам совершенствовать уже имеющиеся модели и строить новые, более точные.

1. Ишемическая болезнь сердца – Российский кардиологический научно-производственный комплекс [Электронный ресурс]. URL: <https://www.rusintervention.ru/пациентам/заболевания/ибс/>

2. Кабаков, Р.И. R в действии. Анализ и визуализация данных в программе R / Р.И. Кабаков; пер. с англ. П.А. Волковой. – М.: ДМК Пресс, 2014. – 588 с.

3. Осложнения коронарографии, риски: основные проблемы после реконструкции сосудов через руку – гематома и прочее [Электронный ресурс]. URL: <http://cardiobook.ru/oslozhneniya-koronarografii/>

4. Официальный сайт проекта R. [Электронный ресурс]: URL: <https://cran.r-project.org/>