

ПРИМЕНЕНИЕ МЕТОДА ЛОГИСТИЧЕСКОЙ РЕГРЕССИИ ДЛЯ ПРОГНОЗИРОВАНИЯ ДЕПРЕССИВНЫХ ПЕРЕЖИВАНИЙ ПРИ НАРУШЕНИИ ПРОЛАКТИНЕРГИЧЕСКОЙ ФУНКЦИИ ГИПОФИЗА

В современной науке активно изучается проблема связи гиперпролактинемии и различных психических состояний[1]. Цель исследования - оценка роли социо-демографических факторов при депрессивных состояниях, ассоциированных с гиперпролактинемией при помощи метода множественной логистической регрессии.

Количество участников в анализе исследования 149. Для прогнозирования депрессии у женщин был проведен анализ методом бинарной многофакторной логистической регрессии. Использование метода последовательного отбора (stepwise) [2] независимых переменных, при котором оценивались следующие факторы: уровень образования, отношение к религии, количество детей, употребление кофе, факт курения, употребление алкоголя, физических нагрузок, наличие повышенного пролактина. В итоговую модель вошли факторы, показавшие значимость при предварительном одномерном анализе.

Суммарные баллы, полученные в результате опроса пациенток по опроснику Бека (SF-12), были перекодированы в систему «0» – отсутствие депрессии, «1» – наличие депрессии, порог брался исходя из эмпирических значений среднего по соматической субшкалы.

Анализ данных выполнен с использованием статистических пакетов и R (версия 3.4.1) и STATISTICA 10.

Итоговое уравнение логистической регрессии имеет вид:

$$y = -1,005 + 0,0017x_1 + 0,856x_2, (1)$$

где x_1 – наличие повышенного пролактина, x_2 – факт курения.

Характеристики коэффициентов представлены в таблице 1.

Были проверены предпосылки построения логистической модели: анализ мультиколлинеарности независимых переменных, наблюдения являлись независимыми согласно дизайну исследования. Статистическая значимость логистической модели $\chi^2_{df=1} = 10,876$, $p > 0,005$.

Таблица 1.

Параметры для коэффициентов множественной логистической регрессии

Переменные	коэффициенты	Стандартная ошибка	χ^2	p-значение	ОШ	95%ДИ
(Intercept)	-1.005	0.558	-8.767	< 2e-16 ***	-	-

value_prl	0.0408759	0.0100014	4.087	4.37e-05 ***	1.001	(0.99, 1.003)
smoking	0.2829378	0.0202619	13.964	< 2e-16 ***	2.06	(1.01, 4.18)

Для оценки качества классификации модели была построена ROC-кривая (receiveroperatingcharacteristic)(Рис.1), вычислена величина площади под ROC-кривой, составляющий 0,655, свидетельствует о среднем качестве модели.

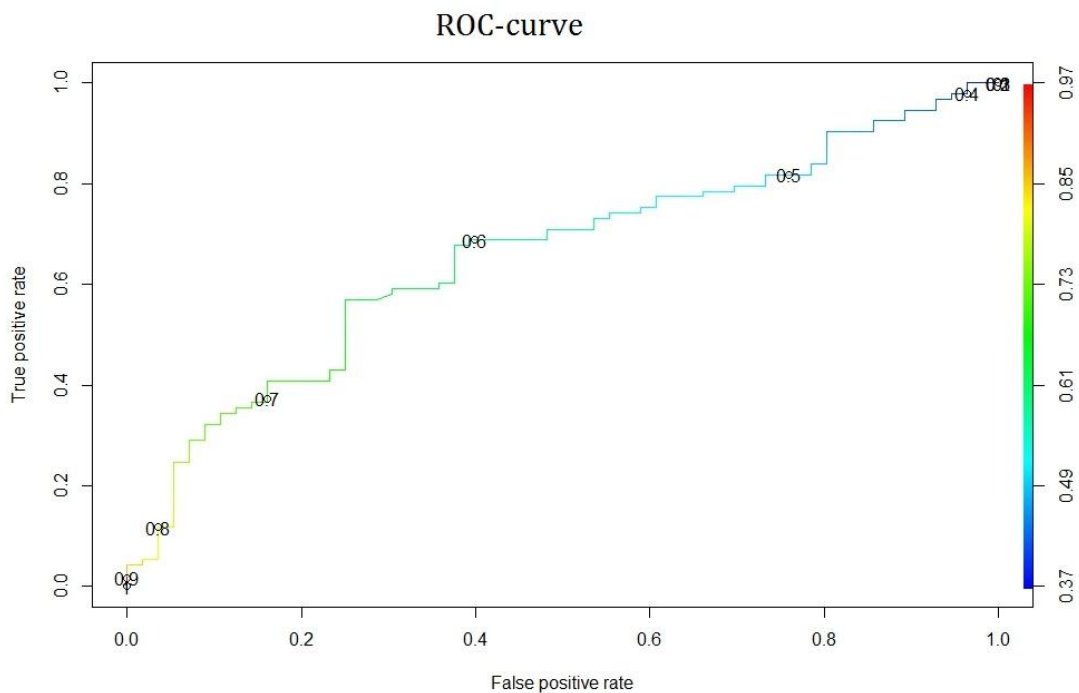


Рис. 1.График ROC-кривой

Все наблюдения с предсказанными значениями меньше или равными 0,617 классифицированы как неудача, остальные - больше 0,617 - как успех (Рис.2). В нашем случае процент правильно предсказанных депрессивных состояний составляет 73%, а отсутствие депрессии – 50%.

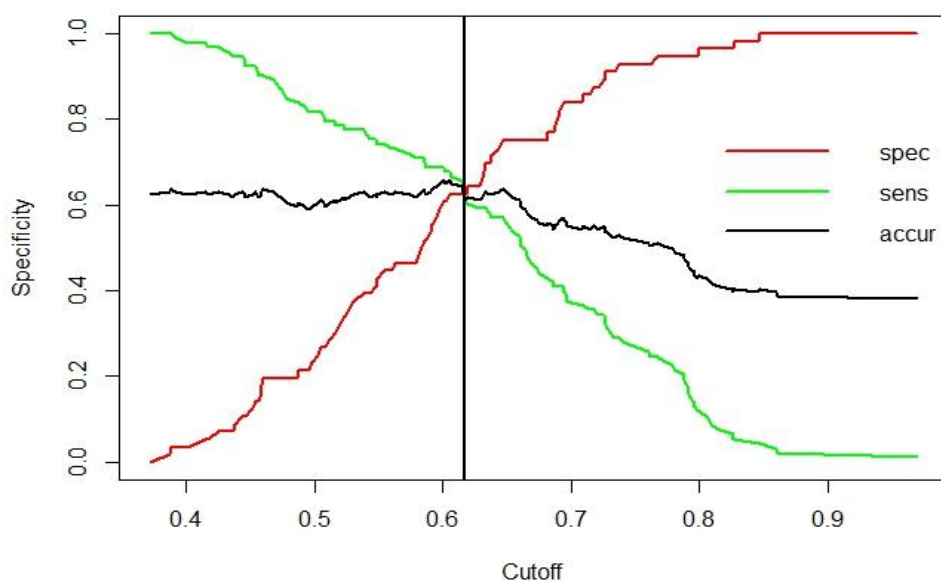


Рис.2. Пересечение трех графиков (специфичность, чувствительность, точность) для определения порога отсечения. График красного цвета – специфичность, зеленого – чувствительность, черного – точность.

При помощи метода множественной логистической регрессии было выявлено, что основными факторами, влияющими на депрессию стали наличие повышенного пролактина и факт курения. По результатам многофакторной логистической модели был получен вклад в развитие депрессии.

Литература:

1. Кузьмин М. Ю., Аталян А. В., Сутурина Л. В. Качество жизни и депрессивные переживания женщин репродуктивного возраста с гиперпролактинемией // Доктор.Ру. 2017. № 13 (142) — 14 (143). С. 52–56.
2. Мастицкий С.Э., Шитиков В.К. (2014) Статистический анализ и визуализация данных с помощью R. – Электронная книга, адрес доступа: <http://r-analytics.blogspot.com>

Научные руководители:

Профессор, зав. кафедры теории вероятности и дискретной математики, д.ф.-м.н. Кузьмин О.В.

Ст.н.с. отдела охраны репродуктивного здоровья ФГБНУ НЦ ПЗСРЧ, к.б.н. Аталян А.В.