

ГЛАВА 4. ПРОГНОЗИРОВАНИЕ И ДИАГНОСТИКА ГИПОТАЛАМИЧЕСКОГО СИНДРОМА И ЕГО РЕПРОДУКТИВНЫХ НАРУШЕНИЙ У ДЕВУШЕК И ЖЕНЩИН

4.1. Результаты использования метода логистической регрессии для прогнозирования и диагностики гипоталамического синдрома и его репродуктивных осложнений у девушек и женщин.

Метод логистической регрессии позволяет оценивать параметры уравнения регрессии, с помощью которого производится прогноз вероятности принадлежности конкретного объекта к тому или иному состоянию. Состояние объекта описывается дискретным качественным признаком. В качестве признаков предикторов выступают номинальные, ранговые или же количественные признаки. Метод был использован для прогноза вероятности принадлежности девушек и женщин к группам больных гипоталамическим синдромом и здоровых (то есть к контрольной группе), а также – к подгруппам с различными вариантами нарушений состояния репродуктивной системы.

В качестве потенциальных предикторов рассматривали как качественные признаки (характеризующие, в основном, анамнез жизни, особенности становления менструальной функции, репродуктивный анамнез), так и количественные признаки (гормональные, биохимические, иммунологические и некоторые другие показатели) (см. Приложение 2). Относительный вклад отдельных предикторов выражали величиной статистики Вальда χ^2 (Wald Chi-Square) а также величиной стандартизованного коэффициента регрессии (Standardized Estimate). Вероятность отнесения отдельного пациента к первой градации вычисляли исходя из уравнения регрессии, которое в общем случае имеет следующий вид:

$$p = \frac{\exp(\beta_0 + \beta_1 * X_1 + \beta_2 * X_2 + \dots + \beta_k * X_k)}{1 + \exp(\beta_0 + \beta_1 * X_1 + \beta_2 * X_2 + \dots + \beta_k * X_k)}$$

где β_i - коэффициенты регрессии для предикторов X_i . Величина β_0 принято называть свободным членом (INTERCEPT)

В качестве критерия согласия реального распределения наблюдений по отдельным градациям признака на основе уравнения логистической регрессии использовался процент правильной переклассификации (Concordant), а также величина коэффициента связи D-Зоммера (Somers'D), отражающий связь фактической частоты и предсказываемой по уравнению логит-регрессии. Общая оценка согласия модели и реальных данных производилась с использованием теста согласия Хосмера-Лемешова (Hosmer and Lemeshow Goodness-of-Fit Test). В данной главе для всех представленных моделей достигнутый уровень значимости для этого критерия $p > 0,05$, т.е. принимается нулевая гипотеза о согласии модели и реальных данных.

При оценке уравнений регрессии использовался метод пошагового включения предикторов, который ранжирует признаки в соответствии с их вкладом в модель. Уровень значимости для включения (исключения) предикторов в уравнение регрессии задавался таким, чтобы достигнутый уровень значимости статистики Вальда χ^2 ($Pr > \text{Chi-Square}$) для каждого предиктора по окончании пошаговой процедуры не превышал 10%. В тех случаях, когда достигнутый уровень значимости для свободного члена β_0 (INTERCEPT) в уравнении регрессии превышал 5%, производилась повторная оценка уравнения регрессии без включения β_0 .

На первом этапе была создана модель гипоталамического синдрома в смешанной группе девушек и женщин, то есть вне зависимости от возраста.

Выражение для вычисления показателя экспоненты для уравнения регрессии представлено ниже:

$$\beta = 14,01 - 2,7 \times V9 - 2,54 \times V10 - 0,59 \times V34 - 0,01 \times V48 + 16,09 \times Otn3$$

$$\text{Concordant} = 94,9\% \quad \text{Somers' D} = 0,898$$

$$R^2 = 0.5024 \quad \text{Max-rescaled } R^2 = 0.7099$$

Где

14,01- свободный член,

V9-перинатальные факторы (0- не были, 1- были)

V10-черепно-мозговые травмы (0- не были, 1- были)

V34-индекс массы тела ($\text{кг}/\text{м}^2$)

V48-пролактин (мМЕ/мл)

Otn3-соотношение «длительность половой жизни/возраст»

Значение « β » вычисляется как сумма произведений коэффициентов на значения соответствующих признаков. Вероятность отнесения обследуемой к первой градации признака (в данном случае к группе здоровых) или «PRED1» вычисляется по формуле:

$$\text{PRED1} = (\text{EXP}(\beta) / (1 + \text{EXP}(\beta))),$$

а вероятность ее отнесения ко второй градации (больные ГС)

$$\text{PRED2} = 1 - \text{PRED1}$$

Когда вероятность P1 более 0,5, обследуемый относится к первой градации (в данном случае – к здоровым).

Для сопоставления между собой предикторов по силе их влияния на прогнозируемый признак (принадлежность к группе больных ГС или здоровых) используется стандартизованный коэффициент регрессии (Standardized Estimate) - чем больше абсолютное значение этой статистики (параметра), тем сильнее он влияет на зависимый признак, в

нашем случае - на вероятность принадлежности к группе больных или здоровых.

В представленной выше модели предикторы распределились по значимости следующим образом (по мере увеличения значимости с учетом стандартизованного коэффициента регрессии):

V34 (индекс массы тела)	-1.543976
OTN3 (соотношение «длительность половой жизни/возраст»)	1.186755
V48 (концентрация пролактина)	- 0.809115
V9 (перинатальные факторы)	- 0.751598,
V10 (черепно-мозговые травмы)	- 0.527523,

Таким образом, в данной модели гипоталамического синдрома наибольшую значимость имеет индекс массы тела, а наименьшую – информация о перенесенной черепно-мозговой травме.

Пример 1. Обследована девушка 18 лет, номер карты 16, V9=1, V10=0; V34=33; V48=123,6; Otn 3=0,15.

Вычисляем значение $\beta = 14,01 - 2,7 \times 1 - 2,54 \times 0 - 0,59 \times 33 - 0,01 \times 123,6 + 16,09 \times 0,15 = -6,7508$

Вероятность отнесения обследованной к первой градации признака (здоровые) вычисляется по формуле $(EXP(\beta)/(1+EXP(\beta)))$. Значение экспоненты определяется с помощью пакета программ Microsoft Excel. В данном случае вероятность отнесения обследованной к первой градации признака $PRED1=0,0012$.

Соответственно, вероятность ее отнесения ко второй градации (больные ГС) $PRED2=1- PRED1=0,9988$. Поскольку вероятность $PRED1$

менее 0,5, а вероятность PRED2 более 0,5, обследованная относится ко второй градации (больные ГС)

Пример 2. Обследована девушка 20 лет, N карты 6004

$V_9=0; V_{10}=0; V_{34}=20; V_{48}=192,7; O_{tn3}=0,05$

Вычисляем значение $\beta=14,01-2,7 \times 0 - 2,54 \times 0 - 0,59 \times 20 - 0,01 \times 192,7 + 16,09 \times 0,05=1,777$

Вероятность отнесения обследованной к первой градации признака (здоровые) вычисляется по формуле $(EXP(\beta))/(1+EXP(\beta))$. Значение показателя для экспоненты определяется с помощью пакета программ Microsoft Excel. В данном случае вероятность отнесения обследованной к первой градации признака $PRED1= 0,8553$. Соответственно, вероятность ее отнесения ко второй градации (больные ГС) $PRED2=1-PRED1=0,1447$. Поскольку вероятность PRED1 более 0,5, а вероятность PRED2 менее 0,5, обследованная относится к первой градации (здоровые)