Ю. И. Пастухова, М. В. Зарубина, М. С. Минаева (Королев, КИУЭС). **Статистическая оценка вероятности выбора карьеры учениками старших классов.**

Ученикам 10–11 классов двух средних школ Пушкинского района Московской области была предложена анкета, в которой содержался ряд вопросов, основным из которых (для проведения исследования) являлся вопрос о намерении продолжить образование в колледже или в ВУЗе. Остальные вопросы анкеты отражали факторы, которые оказывают влияние на принятие решения о продолжении образования. Было выделено $k = 10$ таких факторов: x_1 — средняя оценка по гуманитарным предметам; x_2 — средняя оценка по естественно-научным предметам; x_3 — количество старших братьев или сестер в семье; x_4 — количество младших братьев или сестер в семье; x_5 — уровень дохода семьи; x_6 — наличие высшего образование у родителей; x_7 — количество членов семьи; x_8 — из них работающих; x_9 — возраст родителей; x_{10} — уровень способностей к обучению.

Рассмотрена линейная модель вида $y_i = \sum_{j=1}^{10} b_j x_{ji} + \varepsilon_i = x_i^T B + \varepsilon_i$, $i = 1, 2, \ldots, n$, где

$$y_i = \begin{cases} 1, & \text{если образование предполагается продолжить в ВУЗе,} \\ 0, & \text{если образование предполагается продолжить в колледже,} \end{cases}$$

$\mathbf{P}\left(\varepsilon_i | x_i\right) = 0$, $p_{1i} \equiv \mathbf{P}\left\{y_i = 1 | x_i\right\} = x_i^T B$, $p_{0i} \equiv \mathbf{P}\left\{y_i = 0 | x_i\right\} = 1 - x_i^T B$.

Для оценивания параметров модели применяется метод максимума правдоподобия. В предположении независимости наблюдений y_i строится функция правдоподобия $L(B) = \prod_{i=1}^{n} p_{1i}^{y_i} p_{0i}^{1-y_i}$. Далее,

$$\ln L(B) = \sum_{i=1}^{n} [y_i \ln p_{1i} + (1 - y_i) \ln p_{0i}] = \sum_{i=1}^{n} [y_i \ln F(x_i^T B) + (1 - y_i) \ln(1 - F(x_i^T B))].$$

Здесь $F(x)$ — некоторая удобная функция распределения. В классе дихотомических моделей, к которым относится рассматриваемая модель, часто используют стандартную логистическую функцию распределения $\mathcal{L}(x) = e^x/(1 + e^x)$. Таким образом, оценки $\overline{B}_j$ $(j = 1, 2, \ldots, 10)$ параметров находятся из системы

$$\sum_{i=1}^{n} \left(y_i - \frac{e^{x_i^T \overline{B}_j}}{1 + e^{x_i^T \overline{B}_j}} \right) x_i = 0, \qquad j = 1, 2, \ldots, 10.$$

Полученные оценки $\overline{B}_j$ параметров позволяют оценить вероятность принятия решения в зависимости от значений выбранных факторов x_j. Для решения этой проблемы рассчитаны предельные эффекты коэффициентов факторов при различных условиях. Было также проведено исследование предсказательной способности модели, которая оказалась достаточно высокой по всем используемым для моделей бинарного выбора показателям. Так, например, процент угаданных исходов w для $n = 69$ найден, исходя из результатов моделирования:

Фактические y_i	Предсказанные 0	Предсказанные 1
0	16	4
1	3	46
69	19	50

$$w = \frac{16 + 46}{69} 100\% = 90\%.$$