Инженерный вестник Дона, №1 (2017)
ivdon.ru/ru/magazine/archive/n1y2017/4058

Регрессионная модель успеваемости учебных групп вуза

М.В. Гранков, В.М. Аль-Габри

Донской государственный технический университет, Ростов-на-Дону

Аннотация: В статье рассматриваются подходы к решению задачи повышения успеваемости обучающихся. На базе мультипликативной свертки базовых факторов, влияющие на результаты экзаменов, предложена структура регрессионной модели. Рассмотрены проблемы мультиколлинеарности факторов и «грубых ошибок». Разработана математическая модель и метод построения регрессионных моделей учебных групп. Предложена система классификации групп, в зависимости от классов сопоставляемых им моделей. С помощью предложенного метода была обработана информация из базы данных оценок портала УУ Донского государственного технического университета. Анализ обработанной информации показал, что адекватность порожденных регрессионных моделей выполнялась для 99,7% групп. Показано, что предложенные модель и метод могут быть использованы в системах поддержки образовательного процесса.

Ключевые слова: высшее образование, успеваемость, регрессионный анализ, факторы, выходная переменная, адекватность модели, коэффициент корреляции.

Введение. Одной из проблем постиндустриального общества является очевидная поляризация образовательных организаций на элитные и массовые. Многие исследователи этой проблемы обнаруживают регрессивный тренд у среднего уровня успешности обучающихся по результатам мониторинга различных национальных образовательных систем. Данные исследования опираются на современные методы системного анализа, моделирования, обработки экспериментальных данных и прогнозирования и являются актуальными [1,6,7,8,10]. В настоящей работе рассматривается подход к моделированию и прогнозированию результатов итогового контроля на примере учебных групп студентов.

Цель. Разработка метода прогнозирования результатов сессии, позволяющего для учебной группы определить степень влияния расписания экзаменов на число полученных отрицательных оценок.

Обоснование структуры регрессионной модели успеваемости. Рассмотрим учебную группу как объект, который преобразует внешние воздействия экзаменатора, дисциплины и расписания сессии в значение успеваемости учебной группы, показанную на экзамене. При этом, выходной переменной будет успеваемость группы, которую будем обозначать Y^*. Ранее, доказано, что основными факторами, влияющими на выходную переменную, могут являться: дисциплина экзамена, экзаменатор и номер экзамена в расписании сессии [1]. В качестве числовых значений первых двух факторов выбрана средняя успеваемость по дисциплине (X_s) и экзаменатора (X_t). Значения X_s и X_t будем называть рангами дисциплины (предмета) и преподавателя. Введем следующие обозначения:

$N = \{1, 2, \dots, i, \dots\}$ - множество натуральных чисел (1)

$N_0 = \{0\} \cup$ - расширенное множество натуральных чисел (2)

$I = [0,1]$ - единичный отрезок действительных чисел (3)

$$E_n = \left\{ r_i \middle| (i = 1, 2, \dots n) \wedge (r_i = \frac{n1_i}{n2_i}) \wedge (n1_i \in N_0) \wedge (n2_i \in N) \wedge (n1_i \leq n2_i) \wedge (n \in N) \right\} \quad (4)$$

— множество неотрицательных рациональных чисел, представленных дробями, значение которых не превышает 1.

$$R_+ = \{ r \mid (r \in R) \wedge (r \geq 0) \} \quad (5)$$

— множество неотрицательных действительных чисел, где $R -$ множество действительных чисел.

$$G = \{1, 2, \dots n_g\}, n_g \in N \quad (6)$$

— множество учебных групп, где n_g – число учебных групп.

$$T = \{1, 2, \dots n_t\}, n_t \in N \quad (7)$$

— множество экзаменаторов (преподавателей), где n_t – число экзаменаторов.

$$S = \{1, 2, \dots n_s\}, n_s \in N \quad (8)$$

— множество дисциплин (предметов), где n_s – число дисциплин.

Очевидно, что $X_t \in E_n$ и $X_s \in E_n$. В отношении факторов X_s и X_t классическая регрессионная модель будет иметь следующий вид:

$$Y^* = a + b_1 X_t + b_2 X_s \qquad (9)$$

где b_1 и b_2 – коэффициенты регрессии, а – сдвиг (постоянный член).

Для модели (9) возникает проблема интерпретации значения сдвига [4]. Как видно из (9) средневзвешенная композиция значений факторов X_s и X_t не обеспечивает нулевое значение выходной переменной Y^* при нулевом значении только одного из них, что следует из определения значений этих факторов. В этом смысле мультипликативная композиция факторов соответствует требованию определения. В качестве такой композиции выберем среднее пропорциональное значений рангов экзаменатора и экзамена (10). Значение гибридного фактора, полученное по формуле (10), будем называть рангом экзамена, который принимал экзаменатор t по дисциплине s.

$$X_{ts} = \sqrt{X_t \cdot X_s} \qquad (10)$$

где: X_{ts} – значение ранга экзамена, X_s – значение ранга дисциплины s, X_t – значение ранга экзаменатора t.

$$X_t = \left(\sum_{i=1}^{nt} n1t_i \Big/ \sum_{i=1}^{nt} n2t_i \right) \qquad (11)$$

nt –число экзаменов, которые принял преподаватель t, $n1t_i$ –число отрицательных оценок, которые поставил преподаватель t на экзамене i, $n2t_i$ – общее число оценок, которые должен был поставить преподаватель t .

Аналогично находится ранг дисциплины X_s.

Основываясь на предположении о влиянии расписания экзаменов на их результат [2], представим эту часть регрессионной модели в виде многочлена второго порядка от номера экзамена (X_e) в расписании сессии группы.

$$y(X_e) = b_1 X_e^2 + b_2 X_e + b_3 \qquad (12)$$

Окончательно получим вид регрессионной модели для выходной переменной с помощью мультипликативной свертки гибридной переменной X_{ts} и квадратного трехчлена (12):

$$Y^* = X_{ts}(b_1 X_e^2 + b_2 X_e + b_3), \qquad (13)$$

где: Y^* – оценка успеваемость учебной группы на экзамене, который проводил преподаватель t, по дисциплине s, X_{ts} – ранг экзамена; X_e – номер экзамена в расписании сессии, b_1, b_2, b_3 – коэффициенты регрессии.

Уравнение регрессии (12) преобразуем в линейную форму с помощью вспомогательных переменных $X_1 = X_{ts} \cdot X_e^2$, $X_2 = X_{ts} \cdot X_e$, $X_3 = X_{ts}$.

В результате выполненных преобразований, уравнение регрессии (13) приобретает следующий вид:

$$Y^* = b_1 X_1 + b_2 X_2 + b_3 X_3 \qquad (14)$$

Уравнение в виде (13) будем называть базовым уравнением регрессии, а уравнение в виде (14) – вспомогательным уравнением регрессии.

Математическая модель метода прогнозирования результатов сессии.

Рассмотрим некоторую абстрактную учебную группу и для нее введем следующие обозначения:

$$Ex = \{ex_1, \mathrm{ex}_2, ... ex_{nx}\} \qquad (15)$$

— множество экзаменов, по которому будет построена регрессионная модель успеваемости группы; nx — число экзаменов.

Экзамен ex_i в множестве Ex представляет собой упорядоченное множество:

$$ex_i = (xts_i, \mathrm{y}_i, is_i, ix_i), i = 1, 2, ... nx \qquad (16)$$

где: xts_i—ранг экзамена, y_i—результат экзамена, is_i—номер сессии, ix_i — номер экзамена в сессии.

Введем два подмножества множества натуральных чисел N:

$$N_{kx} = \{k, | (k \in N) \wedge (k \le kx)\} \qquad (17)$$

$$N_4 = \{k,|(k \in N) \wedge (k \le 4)\} \tag{18}$$

Определим отображение λ:

$$\lambda : Ex \to ExX \tag{19}$$

— где отношение $ExX \subset I \times E_n \times N \times N_4 \times R_+ \times R_+$ является расширением отношения Ex. Каждому кортежу $ex_i = (xts_i, \mathrm{y}_i, is_i, ix_i) \,|\, ex_i \in Ex$ соответствует один кортеж $exx_i = (xts_i, \mathrm{y}_i, is_i, ix_i, x1_i, x2_i) \,|\, exx_i \in ExX$, у которого первые четыре элемента совпадают с элементами кортежа $ex_i \in Ex$, а последние два вычисляются по формулам: $x1_i = xts_i \cdot ix_i^2$, $x2_i = xts_i \cdot ix_i$.

Обозначим транспонированные проекции отношения ExX:

$\boldsymbol{X}_1 = (\pi_5(ExX))^T$, $\boldsymbol{X}_2 = (\pi_6(ExX))^T$, $\boldsymbol{X}_3 = (\pi_1(ExX))^T$, $\boldsymbol{Y} = (\pi_2(ExX))^T$.

В этом случае мы получим вектора $\boldsymbol{X}_1, \boldsymbol{X}_2, \boldsymbol{X}_3, \boldsymbol{Y}$, состоящие из значений:

$$\boldsymbol{X}_i = \left(x_{i,1}, x_{i,2}, \ldots, x_{i,nx}\right), i = 1,2,3, \tag{20}$$

$$\boldsymbol{Y} = \left(y_1, y_2, \ldots. y_{nx}\right) \tag{21}$$

Обозначим через A матрицу размером $3 \times nx$, строки которой совпадают с векторами $\boldsymbol{X}_1, \boldsymbol{X}_2, \boldsymbol{X}_3$ соответственно:

$$A = \begin{pmatrix} x_{1,1} & x_{1,2} & \ldots & x_{1,nx} \\ x_{2,1} & x_{2,2} & \ldots & x_{2,nx} \\ x_{3,1} & x_{3,2} & \ldots & x_{3,nx} \end{pmatrix} \tag{22}$$

Проблема мультиколлинеарности факторов. Эта проблема может привести к вычислительным или статистическим трудностям построения и проверки адекватности регрессионной модели. Причиной проблемы является линейная зависимость факторов модели. Выделяют явную и скрытую (стохастическую) мультиколлинеарности [4,5]. Признаком отсутствия у факторов явной мультиколлинеарности является условие [5]: $rang(A) = 3$. Нарушение этого условия вызывается линейной зависимостью между факторами. Чаще мультиколлинеарность проявляется в скрытой форме. Приведем один из признаков скрытой мультиколлинеарности. Найдем средние значения факторов и оценки их дисперсий по формулам (39,40):

$$\overline{X_i} = \frac{1}{nx}\sum_{j=1}^{nx} x_{i,j}, i = 1,2,3; \tag{23}$$

$$SX_i^2 = \frac{1}{nx-1}\sum_{j=1}^{nx}\left(x_{i,j} - \overline{X}_i\right)^2,\ i = 1,2,3. \tag{24}$$

Вычислим парные коэффициенты корреляции между значениями факторов по формуле (25):

$$r_{ij} = \frac{1}{SX_i \cdot SX_j \cdot (n-1)}\sum_{k=1}^{n}\left(X_{ik} - \overline{X}_i\right)\left(X_{jk} - \overline{X}_j\right);\ i, j = 1,2,3 \tag{25}$$

Из вычисленных значений образуем корреляционную матрицу Q:

$$Q = \begin{pmatrix} r_{1,1} & r_{12} & r_{13} \\ r_{21} & r_{22} & r_{23} \\ r_{31} & r_{32} & r_{33} \end{pmatrix} \tag{26}$$

Матрица Q является симметрической. Элементы на главной диагонали равны единице: $r_{ii} = 1, i = 1,2,3$. Вычислим определитель Δ_Q матрицы Q.

Выдвинем нулевую гипотезу: определитель матрицы Q равен нулю:

$$\begin{aligned} &H_0 : \Delta_Q = 0 \\ &H_1 : \Delta_Q \neq 0 \end{aligned} \tag{27}$$

Вычислим статистику (28), которая имеет распределение χ^2 [5]:

$$\gamma = nx - 1 - \frac{1}{6}(2\cdot(\mathrm{p}+1)+5)\lg\Delta_Q\ , \tag{28}$$

где p — число коэффициентов регрессии. В нашем исследовании:

$$\gamma = nx - 1 - \frac{13}{6}\cdot\lg\Delta_Q \tag{29}$$

Для вычисленного по формуле (29) значения, проверим условие:

$$\gamma > \chi_{кр}^2\left(\alpha, nx\cdot(nx-1)/2\right), \tag{30}$$

где: α — уровень значимости, $nx \cdot (nx-1)$ — число степеней свободы, $\chi^2_{кр}$ — критическое значение распределения χ^2. Если условие выполняется, то нулевая гипотеза принимается ($H_0 : \Delta_Q = 0$) и результаты использования регрессионного уравнения (14) очень ненадежны. В противном случае нулевая гипотеза отвергается и принимается альтернативная ($H_1 : \Delta_Q \neq 0$).

Вычисление коэффициентов регрессии. Для нахождения коэффициентов регрессии использовался классический метод наименьших квадратов (МНК). Нахождение оценок коэффициентов уравнения регрессии свелось к решению системы линейных уравнений (31):

$$\begin{cases} b_1 \cdot \sum_{i=1}^{nx} x_{1,i}^2 + b_2 \cdot \sum_{i=1}^{nx} x_{1,i} \cdot x_{2,i} + b_3 \cdot \sum_{i=1}^{nx} x_{1,i} \cdot x_{3,i} = \sum_{i=1}^{nx} y_i \cdot x_{1,i} \\ b_1 \cdot \sum_{i=1}^{nx} x_{1,i} \cdot x_{2,i} + b_2 \cdot \sum_{i=1}^{nx} x_{2,i}^2 + b_3 \cdot \sum_{i=1}^{nx} x_{2,i} \cdot x_{3,i} = \sum_{i=1}^{nx} y_i \cdot x_{2,i} \\ b_1 \cdot \sum_{i=1}^{nx} x_{1,i} \cdot x_{3,i} + b_2 \cdot \sum_{i=1}^{nx} x_{2,i} \cdot x_{3,i} + b_3 \cdot \sum_{i=1}^{nx} x_{3,i}^2 = \sum_{i=1}^{nx} y_i \cdot x_{3,i} \end{cases} \qquad (31)$$

Оценка адекватности регрессионной модели и значимости коэффициентов регрессии. По вычисленным коэффициентам регрессионной модели вычислим вектор статистических оценок успеваемостей рассматриваемой группы на экзаменах:

$$\boldsymbol{Y}^* = \left(y_1^*, y_2^*, \ldots. y_{nx}^* \right) \qquad (32)$$

В формуле (32) i-я координата вектора $\boldsymbol{Y}^*$ находится по формуле (33):

$$y_i^* = b_1 \cdot x_{1,i} + b_2 \cdot x_{2,i} + b_3 \cdot x_{3,i} \qquad (33)$$

где: b_1, b_2, b_3 являются решениями системы (31), значения $x_{1,i}, x_{2,i}, x_{3,i}$ являются координатами векторов $\boldsymbol{X}_1, \boldsymbol{X}_2, \boldsymbol{X}_3$ (20). По координатам векторов $\boldsymbol{Y}, \boldsymbol{Y}^*$ найдем вектор $\boldsymbol{Y}_e$:

$$\boldsymbol{Y}_e = \left(ye_1, ye_2, \ldots. ye_{nx} \right) \qquad (34)$$

$$ye_i = y_i - y_i^* \qquad (35)$$

Для проверки адекватности модели необходимо найти статистические оценки математических ожиданий и дисперсий случайных величин $\boldsymbol{X}_1, \boldsymbol{X}_2, \boldsymbol{X}_3, \boldsymbol{Y}, \boldsymbol{Y}^*, \boldsymbol{Y}_e, b_1, b_2, b_3$, для этого вычислим:

$$\overline{Y} = \frac{1}{nx}\sum_{i=1}^{nx} y_i\,, \qquad (36)$$

где: y_i — координата вектора $\boldsymbol{Y}$.

$$S_Y^2 = \frac{1}{nx-1}\sum_{i=1}^{nx}\left(y_i - \overline{Y}\right)^2, \qquad (37)$$

$$\overline{Y^*} = \frac{1}{nx}\sum_{i=1}^{nx} y_i^*\,, \qquad (38)$$

где: y_i^* — координата вектора $\boldsymbol{Y}^*$.

$$S_{Y^*}^2 = \frac{1}{nx-1}\sum_{i=1}^{nx}\left(y_i^* - \overline{Y^*}\right)^2, \qquad (39)$$

$$S_e^2 = \frac{1}{k}\sum_{i=1}^{nx}\left(y_i - y_i^*\right)^2, \qquad (40)$$

где: k — число степеней свободы. В нашей задаче $k = nx - 4$. Оценки дисперсий значений коэффициентов регрессии можно найти по формулам:

$$Sb_i^2 = \frac{S_e^2}{SX_i^2 \cdot (nx-1)}, i = 1,2,3 \qquad (41)$$

Оценки параметров распределений случайных величин $\boldsymbol{X}_1, \boldsymbol{X}_2, \boldsymbol{X}_3$, входящие в формулы (41), приведены ранее (24).

$$R = \frac{1}{S_Y \cdot S_{Y^*} \cdot (nx-1)}\sum_{i=1}^{nx}\left(y_i^* - \overline{Y^*}\right)\left(y_i - \overline{Y}\right) \qquad (42)$$

— значение коэффициента корреляции между успеваемостью и ее оценкой по регрессионной модели. Проверка адекватности регрессионной модели связана с оценками и доверительными интервалами коэффициентов $\beta_1, \beta_2, \beta_3$, построенными для генеральной совокупности. Выдвинем нулевую гипотезу о том, что все коэффициенты $\beta_1, \beta_2, \beta_3$ равны нулю:

$$\begin{aligned} &H_0 : \beta_1 = \beta_2 = \beta_3 = 0 \\ &H_1 : \exists i \big| (1 \le i \le 3) \wedge (\beta_i \ne 0) \end{aligned} \tag{43}$$

Если гипотеза H_0 отвергается, то принимается гипотеза H_1 . Для проверки H_0 воспользуемся методом, основанном на F - статистике [4]:

$$F = \frac{k}{m}\left(\frac{R^2}{1-R^2}\right), \tag{44}$$

где: $k-$ число степеней свободы, $m-$ число коэффициентов регрессии. В нашей задаче формула (44) приобретает вид:

$$F = \frac{nx-3}{3}\left(\frac{R^2}{1-R^2}\right). \tag{45}$$

Гипотеза H_0 выполняется, если значение F не превышает критическое значение из таблиц распределения Фишера.

$$F \le f_{кр}\left(\alpha, nx, nx-3\right), \tag{46}$$

где: $f_{кр}\left(\alpha, nx, nx-3\right)-$ критическое значение из таблиц распределения Фишера, $\alpha -$ принятый уровень значимости. Если условие (46) не выполняется, то принимается гипотеза H_1. Построенная модель считается адекватной и может быть использована в прогнозировании. Доверительные интервалы для коэффициентов $\beta_1, \beta_2, \beta_3$ определяются неравенствами (47):

$$b_i - t_{кр}(\alpha, nx-4)\cdot Sb_i \le \beta_i \le b_i + t_{кр}(\alpha, nx-4)\cdot Sb_i; i = 1,2,3, \tag{47}$$

где: $b_i -$ значение соответствующего коэффициента регрессии, $Sb_i -$ его стандартное отклонение (41), $t_{кр}(\alpha, nx-4)$ критическое значение критерия Стьюдента, $\alpha -$ принятый уровень значимости. Простой проверкой значимости $i-$го коэффициента является выполнение для него условия (48):

$$(b_i - t_{кр}(\alpha, nx-4)\cdot Sb_i)\cdot(b_i + t_{кр}(\alpha, nx-4)\cdot Sb_i) > 0 \tag{48}$$

Классы регрессионной модели. Предположим, что для b_3 выполняется:

$$b_3 > 0. \tag{49}$$

Рассмотрим поведение функции $y(X_e) = b_1 X_e^2 + b_2 X_e + b_3$ (12) в зависимости от знаков ее коэффициентов b_1, b_2 и параметров $x_0 = -b_1 / (2 \cdot b_2)$, $\mathrm{D} = b_2^2 - 4 \cdot b_1 \cdot b_3$. По поведению функции $y(X_e)$ на интервале $1 \le X_e \le 4$ можно выделить тринадцать классов регрессионных моделей (табл.1).

Таблица 1.

Классы регрессионных моделей в зависимости от поведения функции $y(X_e)$ на интервале $1 \le X_e \le 4$, в предположении о положительности $b_3 > 0$

№ модели	b_1	b_2	x_0	D	**функция** $y(X_e)$
1	<0	>0	$0 < x_0 \le 1.5$	>0	убывает
2	<0	>0	$1.5 < x_0 \le 3.5$	>0	Есть max
3	<0	>0	$x_0 > 3.5$	>0	возрастает
4	<0	≤0	$x_0 \le 0$	>0	убывает
5	>0	<0	$x_0 \le 1.5$	<0	возрастает
6	>0	<0	$1.5 < x_0 \le 3.5$	<0	Есть min
7	>0	<0	$x_0 > 3.5$	<0	убывает
8	>0	≥0	$x_0 \le 0$	<0	возрастает
9	>0	<0	$x_0 > 0$	≥0	убывает
10	>0	≥0	$x_0 \le 0$	≥0	возрастает
11	0	>0	-	-	возрастает
12	0	<0	-	-	убывает
13	0	0	-	-	Не зависит от X_e

Как видно из анализа табл.1, классы 11 и 12 соответствуют вырождению квадратичной зависимости $y(X_e)$ в линейную, а класс 13 соответствует моделям, которые описывают группы, результаты экзаменов которых не зависят от номера экзамена в сессии.

Анализ грубых ошибок. Под «грубой ошибкой», или выбросом [3,9], будем понимать такую координату вектора $ye_i \in \boldsymbol{Y}_e$ (34), которая по абсолютной величине значительно превосходит остальные координаты и существенно отличается от стандартного. Такая «ошибка» может быть связана с необычной комбинацией условий подготовки и проведения экзамена. Все условия, повлиявшие на результат такого экзамена, должны подвергаться анализу сотрудниками служб, связанных с поддержкой образовательного процесса. В нашем исследовании признаком аномального результата экзамена служит выполнение критерия Романовского [3]:

$$|ye_i| \geq t(\alpha,k)\cdot S_e, \tag{50}$$

где: $t(\alpha,k)-$ квантиль двухстороннего распределения Стьюдента, $\alpha-$ принятый уровень значимости, k – число степеней свободы [3].

Алгоритм поиска «грубых ошибок»:

1 ПОСТРОИТЬ первичную регрессионную модель по результатам nx экзаменов, ПРОВЕРИТЬ адекватность модели и значимость ее коэффициентов;
2 ЕСЛИ первичная модель адекватна ТО
3 ПОСТРОИТЬ вектор $\boldsymbol{Y}_e$;
4 В векторе $\boldsymbol{Y}_e$ НАЙТИ $ye_k \,|\, (|ye_k| = max(|ye_i|), \forall i \,|\, i = 1,2,...nx) \wedge (1 \leq k \leq nx)$;
5 БЛОКИРОВАТЬ $k-$ю координату векторов $\boldsymbol{X}_1, \boldsymbol{X}_2, \boldsymbol{X}_3, \boldsymbol{Y}$;
6 БЕЗ учета блокированного экзамена, по результатам $nx-1$ экзаменов, ПОСТРОИТЬ вторичную регрессионную модель, ПРОВЕРИТЬ адекватность модели и значимость ее коэффициентов;
7 ЕСЛИ вторичная модель адекватна ТО
8   ПО вторичной модели ВЫЧИСЛИТЬ S_e и ye_k;
9   ПРОВЕРИТЬ условие (50) для S_e и ye_k с $(nx-1)$ степенями свободы;
10   ЕСЛИ условие (50) выполняется ТО
11   ПРИНЯТЬ вторичную модель как окончательную;
12   ИНАЧЕ
13   ПРИНЯТЬ первичную модель как окончательную;

```
    ВСЕ_ЕСЛИ
  ВСЕ_ЕСЛИ
ВСЕ_ЕСЛИ
```

В случае, если предложенный алгоритм нашел экзамен с «грубой ошибкой», то его можно применить еще раз.

Метод прогнозирования и анализа результатов сессии группы. Для того, чтобы реализовать прогнозирование, необходимо иметь некоторую базовую информацию. Пусть для каждой группы из множества G разработан учебный план, в котором для каждой сессии из базовой информации указаны списки: дисциплин сессии, экзаменаторов, результатов экзаменов.

Алгоритм метода прогнозирования и анализа результатов сессии группы:

```
список_выбросов:=NULL
новые_выбросы:=TRUE
ЦИКЛ_ПОКА новые_выбросы
  новые_выбросы:=FALSE
  ПОСТРОИТЬ ПО НЕ БЛОКИРОВАННЫМ ЭКЗАМЕНАМ X_t И X_s
  ЦИКЛ_ПОКА ЕСТЬ ГРУППЫ
    ПО НОВОЙ ГРУППЕ ПОСТРОИТЬ РЕГРЕССИОННУЮ МОДЕЛЬ
    ЕСЛИ модель адекватна ТО
      НАЙТИ выброс
      ЕСЛИ выброс НАЙДЕН ТО
        ДОБАВИТЬ выброс в список_выбросов
        новые_выбросы:=TRUE
      ВСЕ_ЕСЛИ
    ВСЕ_ЕСЛИ
    ПЕРЕЙТИ К СЛЕДУЮЩЕЙ ГРУППЕ
  ВСЕ_ЦИКЛ
ВСЕ_ЦИКЛ
```

Результаты экспериментальных исследований. Возможности применения разработанного метода прогнозирования успеваемости проверялись на программном стенде [1]. В экспериментах использовались данные о результатах сессий групп, которые учились в 2012/2013 учебном году, для которых в сессиях имеется полная информация об успеваемости, начиная с первого семестра. Всего таких групп оказалось 334. Из них, после применения алгоритмов рассмотренного метода, только у одной группы (0,3%) построенная регрессионная модель оказалась неадекватной. Еще у пяти групп (1,5%) не подтвердилась гипотеза (49). Распределение оставшихся групп по классам моделей, описанных в табл.1, представлены на рис.1.

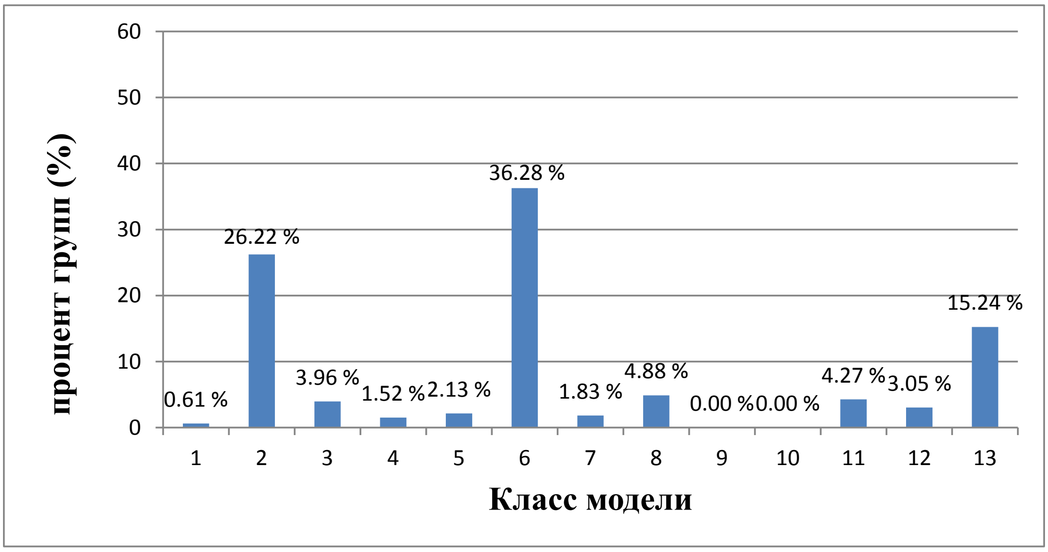

Рисунок 1 – Распределение групп по классам моделей.

Как видно из рис.1, модели более чем 84% групп, участвующих в эксперименте, отражают, в той, или иной степени, зависимость результатов экзаменов от его номера в сессии. Моделям, вырождающимся в линейные, соответствуют 7,32% групп. Процент групп, для которых их модели отражают только нарастающую усталость в процессе сдачи экзаменов, превышает значение 7%. Группы, которые от экзамена к экзамену монотонно адаптируются к процессу их сдачи, повышая результаты, при прочих равных условиях, составили более 15%. Наибольшие проценты групп соответствуют моделям второго и шестого классов. В сумме этим двум моделям соответствуют более 62% групп.

Выводы:

1. Использование структуры регрессионной модели, основанной на мультипликативной свертке факторов, позволило разработать метод прогнозирования результатов итогового контроля в образовательных системах. Применение данного метода для более трехсот групп ДГТУ обеспечило адекватность 99,7% полученных моделей. При этом, значимое влияние расписания на результаты итогового контроля обнаруживается в более, чем 82% моделей, что указывает на возможность использование этого метода в системах принятия решений по управлению успеваемостью.
2. Предложенная система классификации групп, базирующаяся на системе классов их моделей, позволяет выделять множества групп, для которых рационально подбирать последовательность сдачи дисциплин, с целью уменьшения суммарного количества отрицательных оценок за период итогового контроля.

Литература

1. Гранков М.В., Аль-Габри В.М., Горлова М.Ю., Анализ и кластеризация основных факторов, влияющих на успеваемость учебных групп вуза// Инженерный вестник Дона, 2016, №4. URL: ivdon.ru/ru/magazine/archive/n4y2016/3775.

2. Гранков М.В., Аль-Габри В.М. Постановка задачи автоматизации построения расписания экзаменов студентов вуза. Системный анализ, управление и обработка информации: Труды 5-го Международного

семинара (п.Дивноморское 2-6 октября 2014 г.) Под общ. ред. Р.А.Нейдорфа: - Ростов н/Д: - ДГТУ, 2014. — 472с.

3. Румшинский Л.З. Математическая обработка результатов эксперимента. –М.: «Наука», 1971.-192 с.
4. Сигел Э.Ф. Практическая бизнес-статистика. Пер.с англ. - М.: Издательский дом «Вильямс», 2008.- 1056 с.
5. Причины и последствия мультиколлинеарности. URL: metr-ekon.ru/index.php?request=full&id=420
6. Дорохина Е.С., Хорошко А.А., Реализация программы академической и социальной адаптации студентов 1 курса в техническом ВУЗе// Инженерный вестник Дона, 2013, №1. URL: ivdon.ru/ru/magazine/archive/n1y2013/1572.
7. Mehtap Erguven. Comparison of the efficiency of principal component analysis and multiple linear regression to determine students' academic achievement //Application of Information and Communication Technologies (AICT), 2012 6th International Conference, Oct. 2012. pp. 1 – 5.
8. Bekim Fetaji, Xhelal Jashari, Majlinda Fetaji. Devising and evaluating UBT model of student e-Service Information System using regression analyses // Embedded Computing (MECO), 2016 5th Mediterranean Conference. June 2016. pp. 351 – 356.
9. Дрейнер Н., Смит Г. Прикладной регрессионный анализ. Пер.с англ. – 2-е изд., перераб. и доп. - М.: Финансы и статистика, 1986.- 366 с.
10. Пучков Е.В., Пономарева Е.И., Разработка информационно-аналитической системы на основе многомерного хранилища данных// Инженерный вестник Дона, 2012, №4. URL: ivdon.ru/ru/magazine/archive/n4p1y2012/1123

References

1. Grankov M.V., Al'-Gabri V.M., Gorlova M.U., Inženernyj vestnik Dona (Rus), 2016, № 4. URL: ivdon.ru/ru/magazine/archive/n4y2016/3775.
2. Grankov M.V., Al'-Gabri V.M. Postanovka zadachi avtomatizatsii postroyeniya raspisaniya ekzamenov studentov vuza [Problem statement of automating the scheduling of student exams timetable at the university]. Trudy 5-go Mezhdunarodnogo seminara (p.Divnomorskoye 2-6 oktyabrya 2014 g.) Pod obshch. red. R.A.Neydorfa: Rostov n/D: DGTU, 2014. 472p.
3. Rumshinskiy L.Z. Matematicheskaya obrabotka rezul'tatov eksperimenta [The mathematical processing of the experimental results]. M.: «Nauka», 1971. 192 p.
4. Sigel E.F. Prakticheskaya biznes-statistika [Applied Business Statistics]: Per.s angl. M.: Izdatel'skiy dom «Vil'yams», 2008. 1056 p.
5. Prichiny i posledstviya mul'tikollinearnosti [Causes and consequences of multicollinearity]. URL:metr-ekon.ru/index.php?request=full&id=420
6. Dorokhina Ye.S., Khoroshko A.A., Inženernyj vestnik Dona (Rus), 2013, № 1. URL: ivdon.ru/ru/magazine/archive/n1y2013/1572/.
7. Mehtap Erguven. Comparison of the efficiency of principal component analysis and multiple linear regression to determine students' academic achievement. Application of Information and Communication Technologies (AICT), 2012 6th International Conference, Oct. 2012. pp. 1-5.
8. Bekim Fetaji, Xhelal Jashari, Majlinda Fetaji. Devising and evaluating UBT model of student e-Service Information System using regression analyses. Embedded Computing (MECO), 2016 5th Mediterranean Conference. June 2016. pp. 351 – 356.
9. Dreyner N., Smit G. Prikladnoy regressionnyy analiz [Applied regression analysis]. Per.s angl. 2-ye izd., 1986. 366 s.
10. Puchkov EV, Ponomareva EI. Inženernyj vestnik Dona (Rus), 2012, No. 4.

URL: ivdon.ru/en/magazine/archive/n4p1y2012/1123