

В. В. ГАРБАРУК, П. В. ГЕРАСИМЕНКО, С. В. ЕЛИЗАРОВ,
В. А. ХОДАКОВСКИЙ (ПГУПС, Россия)

ОЦЕНКА ЧИСЛЕННОСТИ УЧЕБНО-ВСПОМОГАТЕЛЬНОГО ПЕРСОНАЛА ЛАБОРАТОРИЙ ПО СТАТИСТИЧЕСКИМ ПОКАЗАТЕЛЯМ КАФЕДР

Викладені методичні вказівки щодо побудови математичної моделі оцінювання кількості навчально-допоміжного персоналу для забезпечення учбових практик всіх видів.

Изложены методические указания по построению математической модели оценивания количества учебно-вспомогательного персонала для обеспечения учебных практик всех видов.

The article provides methodical guideline on construction of mathematical model for evaluation of the number of auxiliary educational staff for provision of training workshops of all kinds.

В связи с существенными изменениями связей вузов с внешними организациями, обеспечивавшими учебные практики всех видов, потребовалось изменение внутренней штатной структуры учебных лабораторий, на которые возложены функции по обеспечению учебного процесса, в том числе в той части, которая ранее обеспечивалась внешними организациями.

В статье предлагается методический аппарат по построению математической модели оценивания количества учебно-вспомогательного персонала на основе опыта, определившего структуру и численность лабораторий в вузе. С этой целью на первом этапе надлежало провести анализ штатной структуры существующих лабораторий, которая сложилась до 90-х гг., установить факторы, определившие количество сотрудников лабораторий, и смоделировать зависимости между количеством сотрудников лабораторий и определяющими их число факторами.

Как известно, на количество штатных сотрудников учебных лабораторий влияет большое число факторов. Из числа этих факторов руководители кафедр и преподаватели, прежде всего, выделяют следующие: стоимость материальной базы кафедры, в рублях; количество лабораторий на кафедре; трудозатраты на проведение лабораторных работ, в часах; трудозатраты на проведение расчетно-графических работ, в часах; трудозатраты на проведение курсового и дипломного проектирования (курсовых и дипломных работ), в часах; общий объем занятий, проводимых кафедрой, в часах; объем производственной практики, в часах.

Построение методического аппарата и оценки влияния на штат перечисленных факторов осуществлялись на основе обработки выборки статистических данных по 53 кафедрам ПГУПС

Первоначально в качестве регрессионной зависимости между численностью учебно-вспомогательного персонала и каждой из девяти независимых переменных была выбрана линейная связь. Тогда при наличии девяти объясняющих переменных модель множественной линейной регрессии имеет вид

$$\hat{y}_{x_i}^{(1)} = a + b_1x_1 + b_2x_2 + b_3x_3 + b_4x_4 + b_5x_5 + b_6x_6 + b_7x_7 + b_8x_8 + b_9x_9,$$

где $\hat{y}_{x_i}^{(1)}$ – теоретическое значение численности штата лабораторий, чел.; x_1 – сумма материальных ценностей, тыс. руб.; x_2 – число лабораторий; x_3 – объем лабораторных работ, час; x_4 – объем расчетно-графических работ, час; x_5 – объем курсового проектирования, час; x_6 – объем дипломного проектирования, час; x_7 – объем производственной практики, час; x_8 – общая нагрузка, час; x_9 – численность преподавательского состава, чел.

На основании выполненных расчетов с помощью метода наименьших квадратов получена функция множественной регрессии:

$$\hat{y}_{x_i}^{(1)} = 2,28946 + 0,00003x_1 + 0,38697x_2 + 0,00101x_3 - 0,00067x_4 + 0,00006x_5 + 0,00002x_6 - 0,00172x_7 + 0,00001x_8 + 0,12135x_9.$$

Поскольку в модели множественной регрессии принято 9 объясняющих переменных, то для выявления качества линейной зависимости результата от факторов используется коэффи-

циент множественной детерминации. Он представляет собой отношение суммы квадратов отклонений функции регрессии от среднего значения и полной суммы квадратов отклонений:

$$r_{\hat{y}_x, 1, 2, \dots, k}^2 = \frac{\sum_{i=1}^n (\hat{y}_{x_i} - \bar{y})^2}{\sum_{i=1}^n (y_i - \bar{y})^2},$$

где \hat{y}_{x_i} – теоретическая численность штата; y_i – реальная численность штата; \bar{y} – среднее значение реальной численности штата.

Как известно, для моделей множественной регрессии целесообразно вычислять скорректированный коэффициент детерминации $r_{\hat{y}_x, 1, 2, \dots, k}^2$, чтобы учесть влияние как объясняющих переменных, так и объема выборки. Скорректированный коэффициент множественной детерминации вычисляется по формуле

$$r_{\text{корр}}^2 = 1 - \left[\left(1 - r_{\hat{y}_x, 1, 2, \dots, k}^2 \right) \frac{n-1}{n-k-1} \right],$$

где k – количество независимых переменных в регрессионной модели.

Хотя метод наименьших квадратов позволяет вычислить численность штата с минимальным отклонением от наблюдаемых значений, наличие суммы квадратов ошибок свидетельствует о том, что линейная регрессия не дает абсолютной точности прогноза, если конечно, точки прогноза не лежат на регрессионной прямой. Следовательно, необходима статистика, которая позволила бы оценить отклонение предсказанных значений штата кафедры от его реальных значений. В качестве такой статистики принято использовать среднеквадратическую ошибку результата

$$S_{\hat{y}_x} = \sqrt{\frac{\sum_{i=1}^n (y_i - \hat{y}_{x_i})^2}{n-2}},$$

где \hat{y}_{x_i} – теоретическая численность штата; y_i – реальная численность штата.

В табл. 1 представлены значения коэффициентов множественной корреляции, множественной детерминации, скорректированного коэффициента детерминации и среднеквадратической ошибки рекомендуемого штата для полученной модели.

Таблица 1

Регрессионная статистика

Коэффициент множественной корреляции	Коэффициент множественной детерминации	Скорректированный коэффициент детерминации	Среднеквадратическая ошибка результата
0,736	0,541	0,446	2,986

Из результатов представленных в табл. 1 следует, что 54 % изменения численности штата объясняется изменением всех девяти независимых переменных, а 46 % изменчивости числа работников кафедры определяется другими факторами, не учтенными в регрессионной модели. Причем значение скорректированного коэффициента детерминации равно 0,45 свидетельствует, что 45 % вариации числа работников кафедры объясняется изменчивостью факторов, а также выбранным их количеством и объемом выборки.

Среднеквадратичная ошибка результата равна 2,986. Она указывает на статистически значимую зависимость, существующую между переменными и помогает найти интервальные значения искомой переменной.

На рисунке приведены остатки регрессионной модели, то есть отклонения теоретической численности штата работников от реального числа работников вспомогательного персонала кафедр.

Из анализа графика отклонений, представленного на рисунке видно, что они одинаково часто принимают как положительные, так и отрицательные значения. Это позволяет предполагать, что регрессионные остатки не имеют автокорреляции. Для подтверждения этого предположения используется статистика Дарбина-Уотсона. Эта статистика оценивает корреляцию между соседними отклонениями по соотношению

$$D = \frac{\sum_{i=2}^n (e_i - e_{i-1})^2}{\sum_{i=1}^n e_i^2},$$

где e_i – остаток, соответствующий i -ому уровню.

Если между соседними отклонениями существует положительная автокорреляция, значение D будет близким к нулю. Если остатки не коррелируют между собой, значение D стремится

к 2. В случае отрицательной автокорреляции значение статистики Дарбина-Уотсона близко к 4.

Для рассматриваемой модели табличные нижнее и верхнее критические значения статистики Дарбина-Уотсона соответственно равны 1,22 и 2,78. Расчетное значение D -статистики равно 1,40. Поскольку $1,22 < D < 2,78$, то авто-

корреляция между остатками отсутствует, а следовательно, данный критерий подтверждает вывод, что линейная функция регрессии адекватно описывает опытные данные. Результаты дисперсионного анализа при уровне значимости $\alpha = 0,05$ представлены в табл. 2.

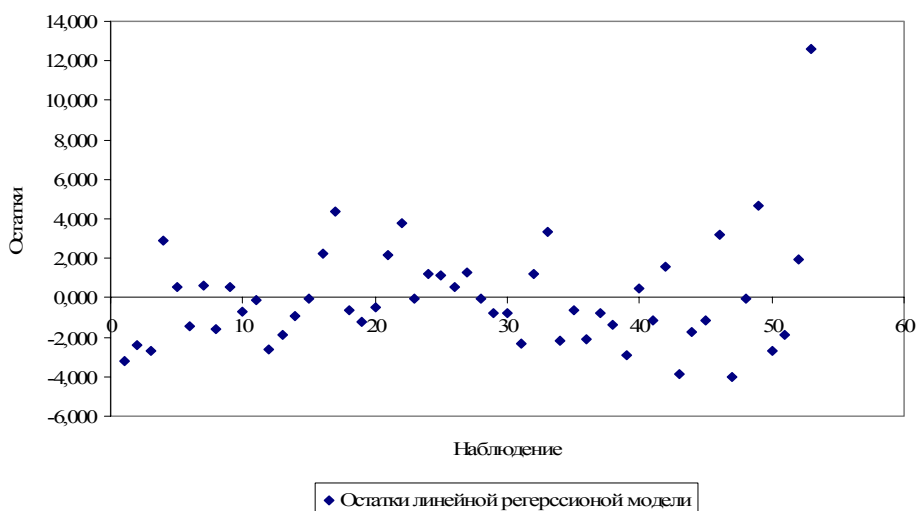


Рис. 1. Отклонения теоретической численности штата от реальной

Таблица 2

Дисперсионный анализ

	Степени свободы	Сумма квадратов	Дисперсия	F -статистика
Регрессия	9	452,617	50,291	5,642
Остаток	43	383,277	8,913	значимость F
Общее	52	835,894		0,000039
Коэффициент a	2,28946	1,15331	1,98512	0,05353
Коэффициент b_1	0,00003	0,00009	0,34373	0,73272
Коэффициент b_2	0,38697	0,21437	1,80515	0,07806
Коэффициент b_3	0,00101	0,00028	3,63468	0,00074
Коэффициент b_4	-0,00067	0,00090	-0,74164	0,46234
Коэффициент b_5	0,00006	0,00032	0,17815	0,85944
Коэффициент b_6	0,00002	0,00047	0,04499	0,96432
Коэффициент b_7	-0,00172	0,00135	-1,26826	0,21153
Коэффициент b_8	0,00001	0,00015	0,06846	0,94574
Коэффициент b_9	0,12135	0,11553	1,05038	0,29941

Для проверки нулевой гипотезы – об отсутствии линейной связи между штатом и объясняющими параметрами – использован F -критерий.

Расчетная F -статистика равна среднему квадрату, обусловленному регрессией, деленному на дисперсию ошибок

$$F = \frac{\frac{1}{k} \sum_{i=1}^n (\hat{y}_{x_i} - \bar{y})^2}{\frac{1}{n-k-1} \sum_{i=1}^n (y_i - \hat{y}_{x_i})^2},$$

где \hat{y}_{x_i} – теоретическая численность штата; y_i – реальная численность штата; \bar{y} – среднее значение реальной численности штат.

Тогда, если $F > F_{k,n-k-1}$ при уровне значимости $\alpha = 0,05$, то нулевая гипотеза отклоняется, в противном же случае нулевая гипотеза принимается. Здесь $F_{k,n-k-1}$ табличное критическое значение F -статистики, k – число переменных регрессионной модели, а n – количество кафедр.

Критическое значение F -статистики с 9 и 43 степенями свободы равно 2,11. Расчетная F -статистика равная 5,64 больше критического 2,11, p -значение равно $0,00004 < 0,05$. Следовательно, нулевая гипотеза отклоняется, и штат

кафедры линейно связан хотя бы с одной из объясняющих переменных.

В дальнейшем по аналогичной схеме выполнено построение линейной множественной функции регрессии с различным числом факторов. С помощью критерия Фишера установлены факторы, которые оказывают наиболее сильное влияние на результат. К таким факторам относятся количество лабораторий, количество часов лабораторных работ, стоимость материальной базы и суммарная нагрузка кафедры. В этом случае функция регрессии имеет вид

$$\hat{y}_{x_i}^{(2)} = 2,13114 + 0,41583x_2 + 0,00101x_3 - 0,00114x_7 + 0,12685x_9.$$

В табл. 3 представлены значения коэффициентов множественной корреляции, множественной детерминации, скорректированного коэффициента детерминации и среднеквадратической ошибки рекомендуемого штата для полученной модели.

Таблица 3

Регрессионная статистика

Коэффициент множественной корреляции	Коэффициент множественной детерминации	Скорректированный коэффициент детерминации	Среднеквадратическая ошибка результата
0,729	0,532	0,493	2,855

Для рассматриваемой модели показано, что расчетное значение D -статистики сохраняет значение равное 1,40 и находится в допустимых пределах, а следовательно, регрессионные остатки являются независимыми. Отсутствие автокорреляции между остатками свидетельствует, что данная линейная функция регрессии сохраняет адекватное описание зависимости.

Критическое значение F -статистики с 4 и 48 степенями свободы равно 2,57. Расчетная F -статистика равна 11,02, т. е. больше критического, которое равняется 2,57. Следовательно, нулевая гипотеза отклоняется, и штат кафедры линейно связан хотя бы с одной из объясняющих переменных.

При уровне значимости равном 0,05 с помощью t -статистики с 47 степенями свободы установлена статистическая значимость коэффициентов функции регрессии.

При поиске более качественной модели рассматривались нелинейные зависимости учебно-вспомогательного персонала от указанных четырех факторов. На основании выполненных расчетов получены функции степенной и показательной множественных регрессий. В частности, степенная функция множественной регрессии имеет вид:

$$\hat{y}_{x_i}^{(3)} = 10^{0,446} \cdot 10^{0,027} \cdot 10^{0,00006x_3} \times 10^{0,00002} \cdot 10^{0,009x_9}.$$

В табл. 4 представлены значения коэффициентов множественной корреляции, множественной детерминации, скорректированного коэффициента детерминации и среднеквадратической ошибки рекомендуемого штата для полученной модели.

Таблица 4

Регрессионная статистика

Коэффициент множественной корреляции	Коэффициент множественной детерминации	Скорректированный коэффициент детерминации	Среднеквадратическая ошибка результата
0,764	0,584	0,550	0,164

Для рассматриваемой модели значение статистики Дарбина-Уотсона, то есть расчетное значение D -статистики равно 1,983. Данный критерий подтверждает вывод, что степенная функция регрессии отражает существующую связь между факторами и результатом. Критическое значение F -статистики с 4 и

48 степенями свободы равно 2,57. Расчетная F -статистика равная 16,869.

На основании выполненных расчетов получена также показательная функция множественной регрессии.

В табл. 5 представлены аналогичные значения коэффициентов для показательной функции.

Таблица 5

Регрессионная статистика

Коэффициент множественной корреляции	Коэффициент множественной детерминации	Скорректированный коэффициент детерминации	Среднеквадратическая ошибка результата
0,739	0,545	0,508	0,171

Для рассматриваемой модели значения статистики Дарбина-Уотсона 1,581, кроме степенной и показательной функций регрессии, оценивались и полиномиальные функции вида:

$$y = b_0 + \sum_{i=1}^4 b_i x_i + \sum_{j=1}^k b_j x_j^n + \varepsilon,$$

где $n = 1, 2, 3$.

Анализ регрессионных функций, устанавливающих зависимость количества сотрудников учебных лабораторий от различных факторов, а также оценки точности и надежности полученных соотношений позволяют сделать следующие выводы.

1. Выполненные корреляционный и регрессионный анализы показали, что факторами, наиболее существенно влияющими на количество сотрудников, в предположении линейной зависимости, являются количество лабораторий, количество часов лабораторных работ, стоимость материальной базы и суммарная нагрузка кафедры.

2. Линейная множественная регрессионная зависимость количества сотрудников от всех включенных кафедрами факторов объясняет только на 51 % эту зависимость (коэффициент

детерминации равен 0,51). Нелинейные зависимости существенно результат не изменили.

3. Линейная множественная корреляционная зависимость количества сотрудников от четырех определяющих факторов (количество лабораторий, количество часов лабораторных работ, стоимость материальной базы и суммарная нагрузка кафедры) практически на такую же величину объясняет зависимость, как и с 9 факторами, что также совпадает с результатом корреляционного анализа.

4. Точечные оценки количества сотрудников, выполненные по построенным функциям регрессии, накрываются достаточно широким доверительным интервалом. Это свидетельствует о том, что при заданной надежности оценок (коэффициент значимости равен 0,05) имеет большую погрешность.

5. Повышение точности может быть достигнуто за счет снижения надежности расчетов или за счет поиска более качественных регрессионных функций и факторов, оказывающих в большей степени влияние на количество сотрудников лабораторий.

Поступила в редколлегию 21.06.2005.